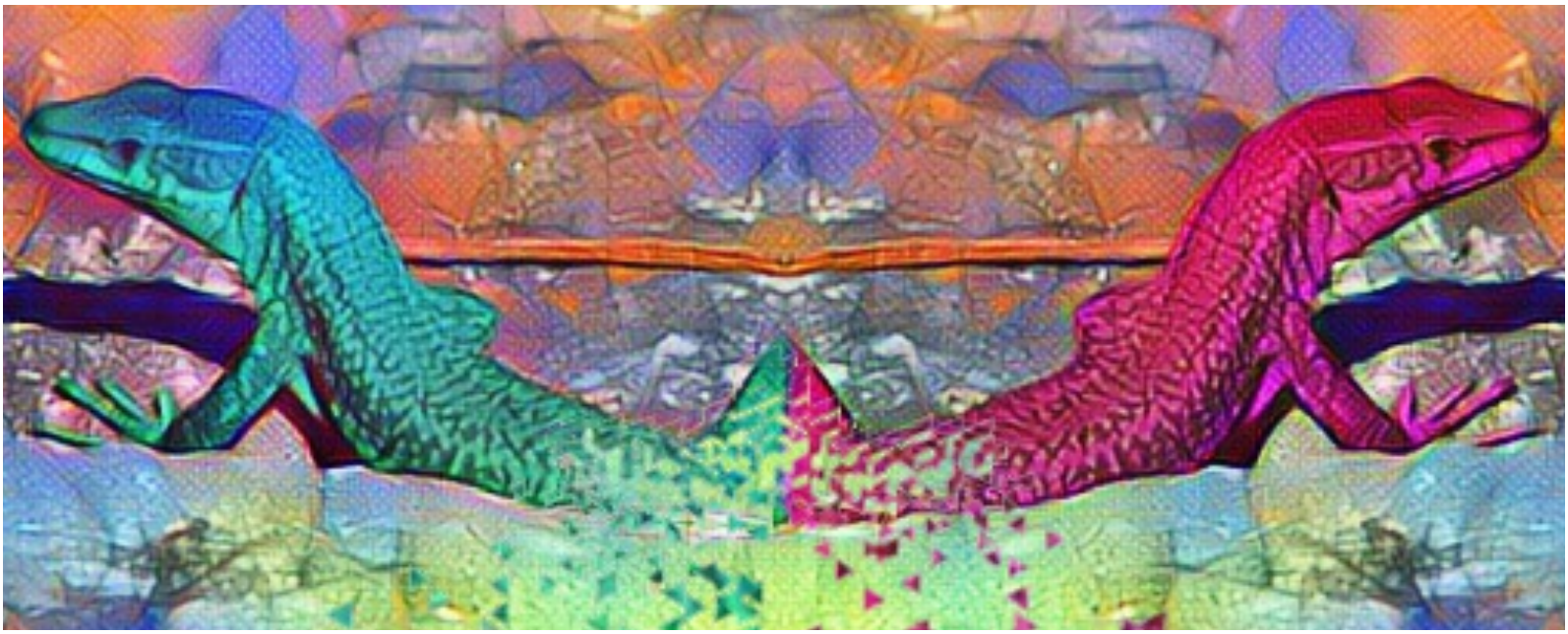

Understanding speciation: a multidisciplinary assessment of hybrid zones using a lizard species complex as model

Guilherme Miguel Caeiro Dias
Doutoramento em Biodiversidade, Genética e Evolução
Departamento de Biologia
2018

Orientadora
Catarina Pinho, PhD, Investigadora Auxiliar
CIBIO/InBIO, Laboratório Associado, Universidade do Porto

Coorientador
Pierre-André Crochet, PhD, Directeur de Recherche
CEFE, CNRS, Université de Montpellier, Université Paul Valéry Montpellier 3

Foreword

In compliance with the no. 2 of article 4 of the General Regulation of Third Cycles of the University of Porto and with the article 31 of the Decree-Law no. 74/2006, of 24 March, with the alteration introduced by the Decree-Law no. 230/2009, of 14 September, the results of already published works were totally used and included in some of the chapters of this dissertation. As these works were performed in collaboration with other authors, the candidate clarifies that he participated in obtaining, interpreting, analyzing and discussing the results in all the works, as well as in the writing of the published forms.

This work was financially supported by Fundação para a Ciência e Tecnologia (FCT, Portugal) through the attribution of a PhD grant (ref.: SFRH/BD/89750/2012), and co-funded by Programa Operacional Potencial Humano (POPH) – Quadro de Referência Estratégico Nacional (QREN), financed by the European Union through the European Social Fund (ESF) and by the Portuguese government.

“We used to make fun of Edgar Anderson by saying that he was finding hybrids under every bush. Then we realized that even the bushes were hybrids.”

Warren H. Wagner

in Abbott *et al.* (2013) Hybridization and Speciation
Journal of Evolutionary Biology, 26, 229–246

Abstract

Speciation is a succession of evolutionary processes prompting divergence between populations that leads to reproductive isolation and to the formation of distinct species. The development of reproductive isolation prevents individuals from diverging populations to successfully reproduce. Most mechanisms of reproductive isolation arise after a long phase or successive phases of divergence in allopatry but may also evolve gradually within a continuous geographic distribution. The study of hybrid zones between two partially reproductively isolated populations has long been recognized as a powerful approach to assess taxonomic status and to study processes involved in speciation and the maintenance of species boundaries. Such regions particularly offer the opportunity to measure the diffusion of genes between diverging taxa, assess the genetic processes responsible for speciation and investigate mechanisms preventing or promoting gene flow. Hybridization may be dependent on the geographical context; therefore, determining species geographic distributions and the underlying causes for species coexistence is of fundamental importance to understand ecological and evolutionary processes involved in differentiation and speciation.

In the first study presented here we took advantage of ecological niche modelling to compare mitochondrial DNA and niche divergence between pairs of lineages belonging to the *Podarcis hispanicus* complex in order to examine patterns of niche divergence and their role in the spatial organization of species. The results highlight the importance of rainfall levels in shaping *Podarcis* wall lizards' distribution. The discordance between significant niche divergence and the patterns of mitochondrial divergence support that genetic divergence across *Podarcis hispanicus* complex most likely occurred in allopatric conditions. Competition after secondary contact is also suggested by partial niche overlap between lineages with strictly parapatric distribution. Complementary suitable niches and competition seem to be two important mechanisms in shaping geographic distributions and restricting the existence of extensive contact zones.

In the other two studies we used restriction site associated DNA sequencing to identify several thousands of informative Single Nucleotide Polymorphisms (SNPs) useful for the study of hybrid zones. In the second study we used a set of thousands of SNP markers to examine the extent of hybridization, level of admixture and variation in

selection against introgression among loci and across the genome between *P. bocagei* and *P. carbonelli*. These two species occasionally mate and hybridize at a narrow region but the large fraction of individuals assigned to one of the parental species clearly show that the hybrid zone is bimodal. This suggests the existence of strong reproductive isolation. Geographic cline analysis corroborates the previous interpretation by demonstrating the presence of strong barriers to gene flow that avoid extensive introgression out of the contact zone. This contrasts with the genomic cline analysis that evidences a large heterogeneity of introgression patterns in the contact zone among loci. About 25% of the genomic regions analysed do not introgress and may be involved in intrinsic barriers to gene flow. The comparison of regions where SNPs were discovered with the *Podarcis muralis* genome denotes that intrinsinc barriers to gene flow are spread across the genome. Notably, the Z chromosome may have a distinct role in reproductive isolation.

The third study was a comparative analysis of contact zones involving *P. carbonelli*, including the same contact zone as in the previous study and three other. First, we used assignment analysis to calculate the proportion of the genome in each individual assigned to parental species. Then, the interspecific heterozygosity was calculated to identify if admixed individuals are first generation hybrids or rather the result of backcrosses and to infer general patterns of interspecific gene flow in the contact zones. Finally, we tested if the proportions of genotypic compositions between all the contact zones were similar. *P. carbonelli* is undoubtably recognized as a species, but our results suggest that, despite the deep divergence times between this and other co-occurring species, reproductive isolation is incomplete between most of them even in the presence of putatively strong pre-zygotic and pos-zygotic isolation mechanisms, i.e. speciation evolved without complete reproductive isolation. Considering that *P. carbonelli* is broadly sympatric with several other species, hybridization may eventually occur anywhere where sympatry exists. Bearing in mind that *P. carbonelli* is listed by IUCN as endangered with a continuous decreasing populational trend and that the few conservation efforts are mainly focused in the habitat protection, we highlight here the urgent need to elaborate a comprehensive conservation plan integrating the effects of hybridization on *P. carbonelli*.

A general outcome from this thesis is that there are strong pre-mating and post-mating mechanisms between several species of *P. hispanicus* complex, preventing extensive interspecific gene flow. Complementary suitable niches and competition seem

to be two important mechanisms in shaping geographic distributions and restricting the existence of extensive contact zones. However, mechanisms of isolation are often incomplete and hybridization occurs. Overall, the results reported in this thesis are consistent with strong, but incomplete, reproductive isolation in the late stages of speciation among species of the *P. hispanicus* complex. Therefore, the effects of hybridization, whether favoring or threatening species persistence, cannot be dissociated from conservation strategies. Moreover, between the two species where we analysed locus specific patterns of introgression we found that the regions potentially involved in reproductive isolation are spread across the genome and are likely to be influenced by a complex combination of mechanisms and selective forces. Future research directions involve applying the same kind of analysis presented here to other contact zones across *P. hispanicus* complex in a more extensive comparative analysis.

Key words: speciation; reproductive isolation; ecological niche modeling; allopatric speciation; niche divergence; geographic clines, genomic clines; hybridization; *Lacertidae*; RAD sequencing.

Resumo

A especiação é uma sucessão de processos evolutivos que conduzem à divergência entre populações, levando por sua vez ao isolamento reprodutivo e à formação de espécies distintas. O desenvolvimento do isolamento reprodutivo impede que os indivíduos de populações divergentes se reproduzam com sucesso. A maioria desses mecanismos surge após uma fase prolongada ou sucessivas fases de divergência em alopatria, mas podem também evoluir gradualmente numa distribuição geográfica contínua. O estudo de zonas híbridas entre duas populações parcialmente isoladas reprodutivamente é uma abordagem eficaz para avaliar o estatuto taxonómico e estudar os processos envolvidos na especiação e manutenção dos limites da espécie. Estas regiões oferecem a oportunidade de medir a difusão de genes entre *taxa* divergentes, avaliar os processos genéticos responsáveis pela especiação e ainda investigar os mecanismos que previnem ou propiciam o fluxo génico. O contexto geográfico onde ocorre a especiação é relevante para os níveis de fluxo genético. Assim, a determinação da distribuição geográfica das espécies e das causas subjacentes a essa distribuição é fundamental para compreender os processos evolutivos envolvidos na diferenciação e especiação.

No primeiro estudo aqui apresentado tirou-se partido da modelação de nicho ecológico para comparar divergência de nicho e divergência mitocondrial entre pares de linhagens pertencentes ao complexo *Podarcis hispanicus*, com o objetivo de examinar padrões de divergência de nicho e o seu papel na organização espacial. Os resultados evidenciam a importância dos níveis de precipitação na distribuição das espécies do complexo. A discordância entre uma divergência de nicho significativa e os padrões de divergência mitocondrial apontam para a ocorrência da diversificação em condições alopátricas no complexo *P. hispanicus*. A sobreposição parcial de nicho entre linhagens com distribuição estritamente parapátrica também sugere competição após contacto secundário. Nichos adequados complementares e competição parecem ser dois importantes mecanismos para a distribuição geográfica e restrição de extensas zonas de contacto.

Nos dois estudos seguintes usou-se um técnica de sequenciação de fragmentos de DNA associados a locais de restrição (RAD *sequencing*, do inglês) para identificar milhares de polimorfismos de nucleotídeos únicos (SNPs, da sigla em inglês)

informativos, úteis para o estudo de zonas híbridas. No segundo estudo, um conjunto de SNPs foi usado para examinar a extensão da hibridação, os níveis de miscigenação e a variação na seleção contra a introgressão entre loci e ao longo do genoma entre *P. bocagei* e *P. carbonelli*. Estas espécies cruzam-se e hibridam ocasionalmente numa região restrita. No entanto, a maior parte dos indivíduos foi identificada como pertencendo a uma das espécies parentais, mostrando claramente que a zona híbrida é bimodal, o que sugere a existência de forte isolamento reprodutivo. Os clinos geográficos demonstram a presença de barreiras ao fluxo génico que impedem a introgressão extensiva fora da zona de contacto. Isto contrasta com a análise de clino genómico, que evidencia uma larga heterogeneidade de padrões de introgressão na zona de contacto entre loci. Estes padrões heterogéneos mostram que a introgressão ocorre mas é restrita à zona de contacto, e podem ser uma consequência de vários mecanismos de natureza intrínseca e extrínseca a cada espécie. A comparação de regiões onde foram descobertos SNPs com o genoma de *P. muralis* denota fortes barreiras ao fluxo génico ao longo do genoma.

O terceiro estudo consistiu numa análise comparativa de zonas de contacto envolvendo *P. carbonelli*, incluindo a zona de contacto já referida no estudo anterior e três zonas adicionais. Numa primeira abordagem calculou-se em cada indivíduo a proporção do genoma atribuído a cada uma das espécies parentais. Posteriormente, calculou-se a heterozigotia interespecífica com o objetivo de identificar se os indivíduos misturados são híbridos de primeira geração ou antes o resultado de retrocruzamentos, e inferir os padrões gerais de fluxo génico interespecífico nas zonas de contacto. Por fim, as proporções das composições genotípicas entre todas as zonas de contacto foram analisadas. *P. carbonelli* é claramente reconhecida como uma espécie distinta; ainda assim, estes resultados sugerem que, apesar dos profundos tempos de divergência, o isolamento reprodutivo é incompleto entre *P. carbonelli* e a maioria das espécies que coocorrem com esta. Isto acontece mesmo na presença de fortes mecanismos de isolamento pré e pós-zigóticos, i.e., a especiação evoluiu sem isolamento reprodutivo completo. Considerando que *P. carbonelli* é amplamente simpátrica com outras espécies, a hibridação pode eventualmente ocorrer em qualquer lugar onde a simpatria exista. Esta espécie está listada pela UICN (União Internacional para a Conservação da Natureza) como ameaçada com uma tendência populacional continuamente decrescente. Sendo que os poucos esforços de conservação estão maioritariamente focados na proteção do habitat, realça-se aqui a necessidade de

elaborar um plano de conservação abrangente, que integre os efeitos da hibridação em *P. carbonelli*.

Em suma, há fortes mecanismos de isolamento pré e pós-acasalamento entre vários pares de espécies do complexo *P. hispanicus*, prevenindo o fluxo génico interespecífico extenso. As distribuições geográficas e a existência de extensas zonas de contacto parecem assentar fortemente em dois importantes mecanismos: nichos complementares adequados e competição. Contudo, os mecanismos de isolamento são por vezes incompletos, ocorrendo hibridação. No geral, os resultados reportados nesta tese são consistentes com um forte, mas incompleto, isolamento reprodutivo nos últimos estádios da especiação no complexo *P. hispanicus*. Assim, os efeitos da hibridação, quer favorecendo ou ameaçando a persistência das espécies, não podem ser dissociados das estratégias de conservação. Adicionalmente, entre as duas espécies em que analisámos padrões específicos de introgressão em determinados loci, a heterogeneidade em tais padrões denota que as regiões potencialmente envolvidas no isolamento reprodutivo estão espalhadas no genoma e são provavelmente influenciadas por uma combinação complexa de vários mecanismos e forças seletivas. As direções futuras de investigação envolvem aplicar o mesmo tipo de análises aqui apresentadas a outras zonas de contacto do complexo *P. hispanicus* numa análise comparativa mais extensa.

Palavras-chave: especiação; isolamento reprodutivo; modelação de nicho ecológico; especiação alopátrica; divergência de nicho; clinos geográficos, clinos genómicos; hibridação; *Lacertidae*; sequenciação por *RAD*.

Index of contents

List of figures

List of tables

Abbreviations

AGU: Aguda

AIC: Akaike information criterion

Alt: altitude

ANOVA: analysis of variance

APre: annual precipitation

AUC: area under the curve

bp: base pairs

CI: confidence interval

COI: cytochrome c oxidase subunit I

DMI: Dobzhansky-Muller incompatibilities

DNA: deoxyribonucleic acid

ENFA: ecological niche factor analysis

ENM: ecological niche model

ESM: Esmoriz

ESP: Espinho

FRA: Francelos

Gb: Giga bases

GBS: genotyping-by-sequencing

GIS: geographical information systems

Het: heterozygosity

HI: hybrid index

IUCN: International Union for Conservation of Nature

JSI: Jaccard's similarity index

LD: linkage disequilibrium

MAD: Madalena

MAF: minor allele frequency

MaxT: maximum temperature of the warmest month

MCMC: Markov chain Monte Carlo

MinT: minimum temperature of the coldest month

mtDNA: mitochondrial deoxyribonucleic acid

Mya: million years ago

ND4: NADH dehydrogenase subunit 4

NGS: next generation sequencing

NRM: non random mating

OD: observed data

PCA: principal component analysis

PCR: polymerase chain reaction

POR: Porto

PreDM: precipitation of the driest month

PreSea: precipitation seasonality

RADseq: restriction site associated DNA sequencing

RM: random mating

RNA: ribonucleic acid

ROC: receiver operating characteristics

SD: standard deviation

SIL: Silvalde

SNP: single nucleotide polymorphism

SS: simulated scenario

SSp: Southern Spain

TMRCA: time to the most recent common ancestor

TOR: Torreira

TSea: temperature seasonality

UPGMA: unweighted pair-group method with arithmetic mean

Chapter 1. General Introduction

1.1. Speciation

Speciation is a series of evolutionary processes triggering divergence between populations and leading to the establishment of distinct species (Figure 1.1.). The definition of a species has been a long-standing debate among biologists and all the attempts to answer the question have always been incomplete in their accounts of biological diversity (Mayden 1997; Hey 2001, 2006). However, the studies on speciation have been focused on the development of reproductive isolation, which follows the biological species concept proposed by Mayr (1942) and defined species as “a group of interbreeding natural populations that are reproductively isolated from other such groups”.

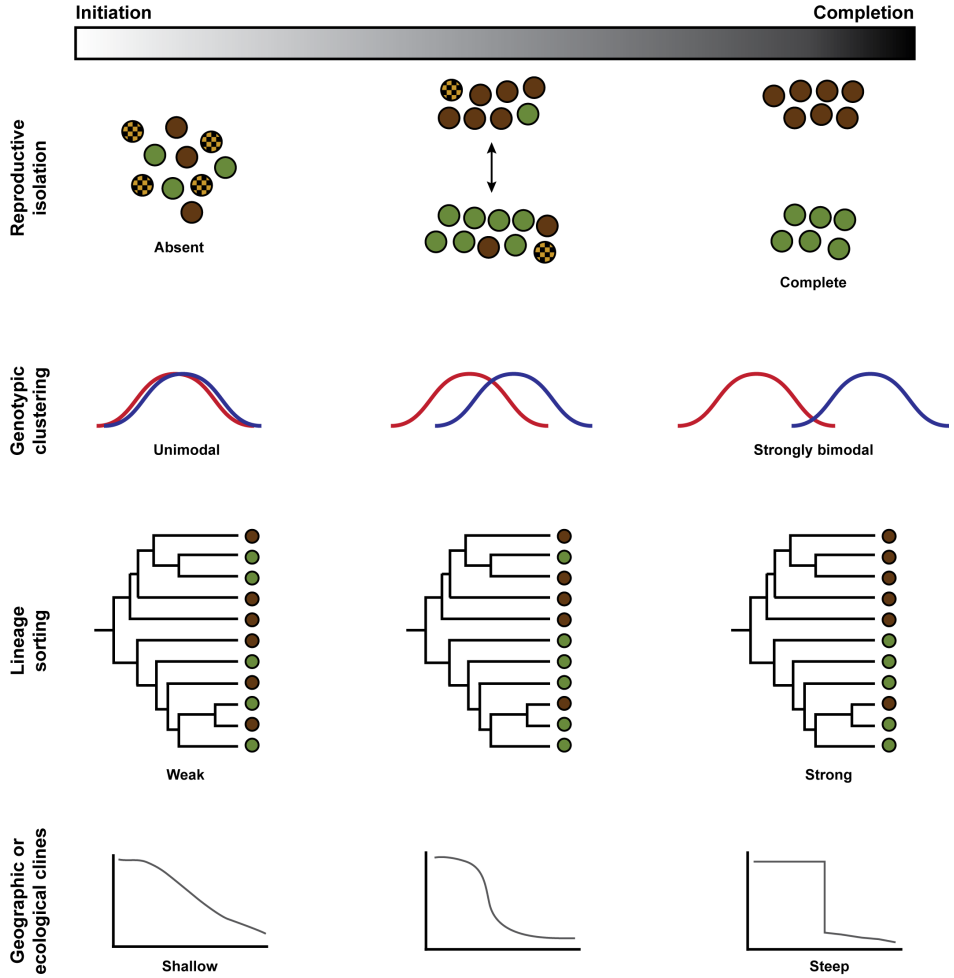

Figure 1.1. Some measures of divergence during the progression of speciation (adapted from Nosil *et al.* 2009). Reproductive isolation is absent when divergence is low and complete when speciation is complete. Distribution of gene frequencies in individuals between two diverging populations vary from unimodal to strongly bimodal. Lineage sorting can vary from weak to strong. Geographic or ecological clines in gene frequency vary from shallow clines in initial stages of speciation to steep clines in the later stages.

The development of reproductive isolation may be related with several mechanisms preventing individuals from diverging populations to produce offspring. The most obvious are ecological factors that impair individuals from distinct populations to meet (e.g. physical barriers, distinct habitats) or phenological factors that prevent mating to happen (e.g. distinct reproductive seasons, distinct times to reach sexual maturity). Sexual isolation and mechanical incompatibilities in sexual structures also avoid the mating to occur. If mating occurs, gametic incompatibilities may also prevent fertilization. These mechanisms that act before the formation of the zygote are the so-called prezygotic reproductive barriers. After the fertilization a number of mechanisms may prevent successful reproduction between divergent populations, i.e. post-zygotic reproductive barriers. Although the zygote may be formed, it is either not viable and does not continue the embryogenesis or undergoes an abnormal development leading embryo death. If the embryo develops successfully it may happen that genic incompatibilities between two or more loci promote low hybrid fitness, or hybrids may have a normal development and viability but are sterile and unable to reproduce.

1.1.1. Modes of speciation

Many studies associate natural selection to the origin of new species, but its role in the speciation process is still an issue under debate. Usually the mechanisms by which selection leads to speciation fall into two general classes: ecological and mutation-order speciation (Schluter 2009). Ecological speciation is the evolution of reproductive isolation between allopatric populations or subgroups within a single population as a result of ecologically mediated divergent selection, i.e. by adaptation to distinct ecological conditions leading to the fixation of alternative alleles in distinct environments (Schluter 2000, 2001; Rundle & Nosil 2005). Mutation-order speciation is the evolution of reproductive isolation by fixation of different random mutations in distinct populations (Mani & Clarke 1990; Schluter 2009). The fixation of alternative alleles may be reached due to intrinsic selective pressures or alternatively by neutral processes, like genetic drift. Speciation under neutral evolution is difficult to occur in the presence of gene flow because beneficial mutations will spread among populations preventing divergence; however, ecological speciation can occur both in presence or absence of gene flow, despite being facilitated in the absence of gene flow (Schluter 2009).

1.1.2. Biogeographical patterns of speciation

Most speciation events are explained by long phase(s) of divergence in allopatry. However, the idea that reproductive isolation may evolve gradually within a continuous geographic distribution, without an abrupt establishment of geographical barriers (Mayr 1963; Kondrashov 1983a; b) has becoming increasingly accepted (Via 2001; Mallet 2008; Nosil 2008), though difficult to demonstrate empirically due to confounding effects of gene flow and recent divergence on genetic differentiation. For instance, weak genetic differentiation between two populations may be due to recent divergence, gene flow or both factors (Nosil 2008). The development of analytical methods like coalescent-based analysis (e.g. Hey *et al.* 2004; Excoffier *et al.* 2013) helped to overcome some of these difficulties.

The geographical context where speciation occurs is relevant to the levels of gene flow between populations undergoing divergence (Mayr 1963). If speciation occurs between geographically isolated populations, i.e. allopatric speciation, gene flow is totally absent. Divergence is thus the result of genetic drift or selection over favored mutations. Over time the acomulation of such mutations contribute to the evolution of reproductive isolation. On the other hand sympatric speciation arises when divergence occurs in the same geographic context due to divergent selection mediated by environmental differences within the same geographical range (Rundle & Nosil 2005). A number of circumstances with partial physical restrictions to gene flow may also occur, like parapatric speciation or speciation through geographical isolation with migration. If populations are not completely separated and gene flow is ongoing, divergence occurs due to selection against gene flow, mediated by any parameter that increases the efficacy of selection (Lenormand 2002) like environmental differences between the populations, assortative mating between locally adapted genotypes, temporal variation in life cycle traits affecting reproduction or low hybrid fitness.

1.1.3. Genetic basis of speciation

Divergence can be explained as a consequence of population genetic processes such as natural selection, sexual selection or genetic drift that may lead to the development of reproductive isolation. Genes that are directly responsible for the evolution of

reproductive isolation (Rieseberg & Blackman 2010; Nosil & Schluter 2011) are defined as speciation genes (Coyne 1992; Wu & Ting 2004; Butlin & Ritchie 2009). Although the number of candidate speciation genes already identified has been growing, especially due to recent advances in genomic technologies and data analysis, only few genes were identified with confidence as speciation genes (Nosil & Schluter 2011).

Genic incompatibilities between two or more loci (epistatic loci) have been long supported as the main reproductive isolation mechanism impeding gene flow between diverging populations (Bateson 1909; Dobzhansky 1937; Muller 1942). The epistatic effects of alleles from one genomic background in an alien genomic background, causing hybrid inviability or sterility, characterize postzygotic reproductive isolation but may also promote prezygotic mechanisms of isolation (e.g sexual isolation, Coyne & Orr 1998).

Besides epistatic interactions, other alternative or complementary genetic mechanisms may drive speciation. Polyploidization is a well-known class of mutations at the origin of immediate speciation both in plants and animals. It can arise through genomic doubling, gametic nonreduction or polyspermy and is at the origin of an abrupt reproductive isolation from the parental species (Otto & Whitton 2000). Chromosomal rearrangements have been also proposed to trigger speciation by enhancing combinations of speciation genes and increasing the effects of reproductive isolation and thus reducing gene flow (Rieseberg 2001; Navarro & Barton 2003). Another recognized phenomenon leading to speciation is the cytoplasmatic incompatibility. In many insects, endosymbionts such as bacteria of the genus *Wolbachia*, induce gamete incompatibility between the sperm of infected males and the eggs of uninfected females establishing a prezygotic barrier (Kambhampati *et al.* 1993; Marshall 2004; Telschow *et al.* 2007).

1.2. Drivers of species distribution

All currently known species are absent from most regions on Earth. Even considering the most ubiquitous species, they do not reach all types of environments due to several limitations. Therefore, determining the limits of species geographic ranges and the underlying causes of their distribution is of fundamental importance in many research fields, including ecology, physiology or evolution.

1.2.1. Determinants of species occurrence

The geographic distribution of species is the result of the interaction of three main groups of factors (Figure 1.2.; Guisan & Thuiller 2005; Soberón & Peterson 2005; Soberón 2007): (i) intrinsic dispersal limitations of organisms, due to their own movements or external mediators, geographical barriers and historical constraints; (ii) spatial distribution of environmental conditions (abiotic factors) favorable to the individuals' establishment, survivorship and reproduction, which constitutes the broad limits of the species geographical distributions; (iii) interactions with other organisms (competitors, predators, preys, pathogens, symbionts) that, together with the availability and dynamics of resources (biotic factors), determine the fine-scale structure of distributions and may change the limits imposed by abiotic factors. Species can thus live in environmentally favorable regions where they are able to disperse and from where they are not excluded by biotic interactions (Soberón 2007; Barbosa *et al.* 2012).

Geographic range limits can be formed in the presence of hard boundaries or environmental gradients, with or without biotic interactions and usually constitute areas of lower densities, lower genetic variation, lower fitness, and higher mortality (Gaston 2009), with some exceptions (Garner *et al.* 2004; Gapare *et al.* 2005).

Changes in species' distribution are an important consequence of modifications in the environment, particularly climate changes, with significant implications for the establishment of contact zones between diverging taxa. Over geological time, climate has suffered large oscillations leading to changes in species' distribution (e.g. during the Pleistocene, climatic fluctuations have generally shaped the diversity and distribution of contemporary organisms; Avise 1998; Avise *et al.* 1998; Gómez & Lunt 2007). Thus, factors defining current patterns of spatial organization also shape the co-occurrence of diverging taxa in the same geographical region and the potential for the establishment of hybrid zones.

1.2.2. Ecological niche

Several definitions of ecological niche have been suggested over time. Grinnell (1917) defined the ecological niche as a portion of the habitat containing the environmental conditions that allow the individuals of a species to survive and reproduce. Elton (1927) further added the function of a species in a community to the ecological niche definition.

Hutchinson (1957) defined the fundamental niche as the subdivision of the ecological niche in *n* environmental variables within which a species can maintain a viable population and persist over time and the realized niche as a part of the fundamental niche where the species is not excluded by competition. Jackson and Overpeck (2000) proposed the concept of potential niche as the part of the fundamental niche that is in fact available for the individuals of a species. Pearson (2007) introduced the concept of occupied niche, to which species distributions are limited by historical, geographic, and biotic factors. Still, some populations may occupy unsuitable habitats due to immigration from nearby populations occupying suitable regions (source-sink dynamics). In these cases the realized distribution goes beyond the fundamental niche, as the species occupies habitats that are not contained in the niche (Pulliam 1988, 2000). On the other hand a species can be absent from suitable habitats for historical reasons or due to limitations in its ability to disperse to those habitats (Holt 2003).

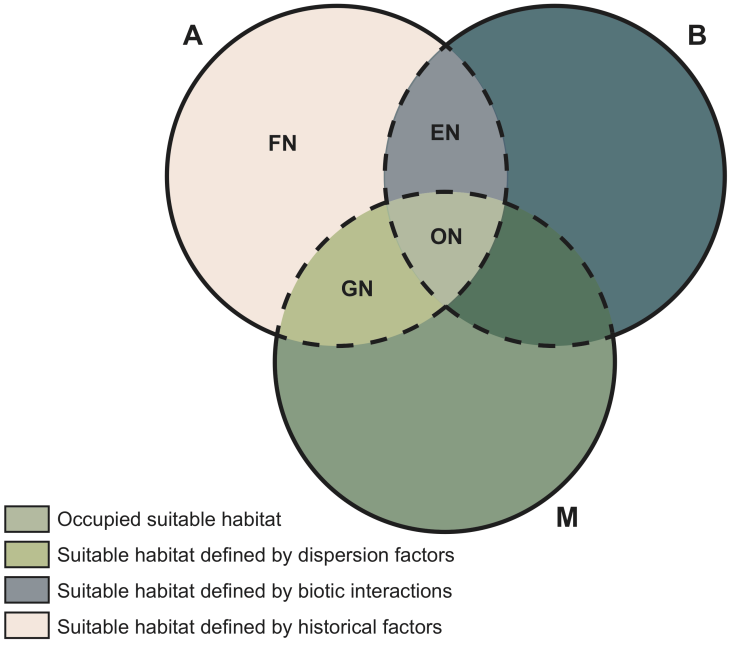

Figure 1.2. Factors determining species occurrence (adapted from Soberón & Peterson 2005). "A" represents the regions with the appropriate set of abiotic factors for the species, and may be refered as the geographic expression of the Hutchinson's fundamental niche (FN). "B" is the region representing the combination of interacting species. "M" is composed by the regions accessible to the species, without barriers to movement and colonization. EN represents the Elton's realized niche and GN Grinnel's niche. The occupied niche (ON) is the suitable habitat and corresponds to the common area to all three factors.

1.2.3. Ecological niche modelling

The analysis of the relationships between species and the environment is a central topic in ecology and the quantification of such relationships is the central issue of predictive geographical modelling (Guisan & Zimmermann 2000). Ecological niche models (ENM) are empirical or mathematical approximations to the ecological niche of a species,

relating the distribution of populations or species with different types of ecogeographical variables (e.g. environmental, topographical, human) in order to identify the limiting factors (Guisan & Zimmermann 2000; Guisan & Thuiller 2005; Sillero 2011). The application of ENM together with Geographical Information Systems (GIS) allows the development of robust models that relate with variables in a geographic context (Guisan & Zimmermann 2000; Sillero 2011; Warren 2012, 2013).

ENMs can be classified as mechanistic and correlative. Mechanistic models are centered on the relationships of variables with a hypothetical direct effect on the species' survival, such as temperature or humidity, and the species' distribution (Barbosa *et al.* 2012). Correlative models are based on statistical correlations between species occurrence and variables that do not necessarily have a direct effect on the species, such as altitude or latitude, but that summarize the effects of various direct factors, and are easier to measure (Guisan and Zimmermann, 2000).

Niche modelling approaches can thus provide valuable information about the environmental conditions that limit species distributions and that have direct consequences for the existence of regions of sympatry between diverging taxa and thus for the potential to establish hybrid zones. Situations in which species contact at the limits of their distributions, following a clear spatial axis, may correspond to areas of gradual environmental transition and/or with biotic interactions. For species with distribution ranges of broad sympatry, environmental conditions should be the main cause of extensive range overlap. More complicated scenarios may be related with patchy contact zones (e.g. association with fragmented habitats, historical factors leading to isolation). Hence abiotic and biotic factors play a major role in the local distribution and population dynamics of the species in contact and the responses of each species to the variations in these factors may allow their coexistence in some regions of sympatry (Barbosa *et al.* 2012).

1.3. Hybrid zones

Hybrid zones are geographic regions where divergent populations meet, mate and successfully reproduce originating viable and frequently fertile offspring (Barton & Hewitt 1985, 1989; Harrison 1990, 1993). Reproduction between diverged individuals from populations that have undergone processes of divergence originating individuals of

mixed ancestry, i.e. hybrids, is referred to as hybridization (Barton & Hewitt 1985, 1989; Harrison 1993). Hybrid zones contain individuals with heterozygous genotypes and loci with intermediate allele frequencies (Barton & Hewitt 1985; Szymura & Barton 1986, 1991). When hybrid offspring is fertile, hybridization my lead to gene flow between divergent populations and together with other processes, such as natural selection and genetic drift, shapes how populations evolve.

1.3.1. Formation and dynamics of hybrid zones

Two main scenarios for the formation of hybrid zones have been proposed. Hybrid zones may be formed between allopatric diverging populations that experience secondary contact (Mayr 1942) or as a result of primary divergence in sympatry by differential environmental selection (Endler 1977; Barton & Hewitt 1985). However, in nature, more complex scenarios are likely to occur, and it is becoming increasingly accepted that hybrid zones may be shaped by the interaction of both primary divergence and secondary contact, possibly through multiple events over time (Gompert *et al.* 2017).

The different contexts in which hybrid zones occur determine the transition from one species to another. Generally hybrid zones encompass a series of gradients in phenotypic traits or allele frequencies across the geographical space that can be highly variable, from a narrow to a wide region where hybridization and/or introgression occur and may be or not related with ecological gradients (Gompert *et al.* 2017). Less frequently, instead of a continuous gradient, hybrid zones can be regions with patchy distributions of populations with hybrids, which are referred to as mosaic hybrid zones (Rand & Harrison 1989).

Hybrid zones can persist for long periods of time. Nonetheless, contemporary hybrid zones may have a recent origin, as a result of human mediated introductions of exotic species or changes in distribution range. Given the changes that species' distributions have undergone due to climate fluctuations, the dynamic nature of many habitats and the changes mediated by anthropogenic activities, it is likely that the composition of contemporary hybrid zones can reflect both recent and historical hybridization events in the genomic composition of populations of interacting taxa (Gompert *et al.* 2017).

Theoretical and empirical studies suggest that hybrid zones are often not stable; they are rather dynamic and can move over time due to a variety of causes. Responses to changes in environmental conditions may drive hybrid zones to move (Britch *et al.* 2001; Taylor *et al.* 2014). Tension zones can shift toward areas of low population density and through directional introgression between parental taxa (Barton 1979; Barton & Hewitt 1985). Differential adaptation of the parents may also cause movement in the direction of the parental taxa with lower fitness (Key 1968; Barton & Hewitt 1985; Bull & Burzacott 2001; While *et al.* 2015).

1.3.2. Analysis of hybrid zones

In order to study speciation through analysis of hybrid zones, a variety of methods might be applied, according to the processes that hybrid zones have been exposed to. Traditionally, studies on speciation have followed two major approaches: interspecific crosses controlled in laboratory (Coyne & Orr 1989) and the study of natural hybrid zones (Szymura & Barton 1986, 1991). Even though laboratory crosses allow a direct quantification of hybrid viability and fertility, the study of hybrid zones between two partially reproductively isolated populations has long been recognized as a powerful approach to assess taxonomic status and to study processes involved in speciation and the maintenance of species boundaries (Barton & Hewitt 1985; Hewitt 1988, 2001; Harrison & Larson 2016). Such regions offer the opportunity to measure the diffusion of genes between diverging taxa, assess the genetic processes responsible for speciation and investigate mechanisms preventing or promoting gene flow (Sequeira *et al.* 2005; Gay *et al.* 2008; Tarroso *et al.* 2014).

Although some studies have documented concordant patterns of introgression (e.g. Zbawicka *et al.* 2014), hybrid zones have frequently variable patterns of introgression that can be identified with few loci (Wilding *et al.* 2001; Payseur *et al.* 2004; Chatfield *et al.* 2010) and genome-wide analysis (Parchman *et al.* 2013; Baldassarre *et al.* 2014; Le Moan *et al.* 2016). Differential patterns of introgression may be a consequence of differential selection among loci e.g. some loci may be under positive selection or associated with reproductive isolation. Both direct selection on loci or combination of loci responsible for the low fitness of hybrids and indirect selection from linkage disequilibrium (see section 1.3.2.1.) may be responsible for differential

introgression in hybrid zones (Gompert *et al.* 2017). Also stochastic processes, like genetic drift, can produce restricted or enhanced introgression patterns (Fitzpatrick *et al.* 2009; Gompert *et al.* 2012).

1.3.2.1. Patterns of introgression in a geographic context

Most hybrid zones are maintained as a result of an equilibrium between migration from adjacent parental populations to the contact zone and selection against hybrids linked to both environmental gradients or intrinsic reduction in fitness of hybrids, referred to as tension zones (Key 1968; Barton & Hewitt 1985, 1989). The development of cline theory provided a methodological background to establish relationships between the distribution of allele frequencies in space and evolutionary processes (Bazykin 1969; Slatkin 1973; Endler 1977; Barton & Hewitt 1985, 1989) by quantifying patterns of gene flow along a geographic gradient crossing hybrid zones and to estimate the strength of selection on individual loci, the overall barrier to gene flow between taxa, and the number of loci contributing to reproductive isolation (Barton & Hewitt 1981; Szymura & Barton 1986, 1991; Mallet *et al.* 1990).

Several models where selection preserves a spatial cline in allele frequencies generate a sigmoidal cline shape (Barton & Gale 1993). The balance between parental migration into the hybrid zone, i.e. gene flow (σ), and the strength of selection against hybrid genotypes (s) determine the cline width (w) in equilibrium. The state of equilibrium between selection and migration can take thousands of generations to reach and may change with environmental fluctuations (Baird 1995). This means that, depending on their age, current hybrid zones may oy may not have reached that state.

The steeper is the cline, the stronger is the selection against gene flow and narrower is the hybrid zone (Mallet *et al.* 1990). If most clines are maintained by a balance between dispersal and similar selective pressures against intermediate genotypes, they are predicted to have similar geographic cline centres and widths, whereas those subject to positive selection on one side of the hybrid zone will have cline centres shifted from the majority of other clines (Barton & Hewitt 1985). However, clines of linked loci will behave similarly. Within a population, it is common that alleles from distinct loci form combinations, even if they are not physically linked, which occur and are inherited together more frequently than expected by random association. This non-

random association of alleles from different loci is referred to as linkage disequilibrium (LD). Commonly LD between genetically divergent populations is caused by migration of parental genotypes to the centre of the hybrid zone. When the selection is strong and occurs on multiple loci, LD creates strong barriers to gene flow, increasing selection (Barton 1983, Barton & Bengtsson 1986).

1.3.2.2. Patterns of introgression in a genomic context

With the recent advances in obtaining genomic data and with the availability of new analytical methods, analysing patterns of differential introgression among loci became easier, allowing the identification of genomic regions contributing to reproductive isolation or that introgress adaptively (Payseur 2010; Harrison & Larson 2016). Genomic cline approaches allow to quantify locus-specific introgression and relate variation in allele frequency comparatively to the genome wide ancestry of introgressed individuals (Gompert & Buerkle 2009, 2011a; Fitzpatrick 2013). Also methods that analyse heterogeneous patterns of ancestry across the genome from multiple populations allow the reconstruction of past gene flow dynamics and the identification of genomic regions acting against gene flow (Gravel 2012; Sedghifar *et al.* 2016).

Several genomic cline models have been proposed to analyse genomic patterns of introgression, such as the concordance model (Szymura & Barton 1986), the regression-based models (Lexer *et al.* 2007; Gompert & Buerkle 2009), the logit-logistic model (Fitzpatrick 2013) or the Bayesian genomic cline model (Gompert & Buerkle 2011b). Genomic cline methods relate introgression with an admixture gradient that can be applied in hybrid zones without considering a spatial gradient. Such methods are thus appropriate to study introgressive hybridization both along a geographic transect and where it is not possible to establish a clear spatial axis as mosaic hybrid zones or in the extreme cases where isolated populations are surrounded by other divergent populations. Bayesian genomic cline model estimates individual ancestry summarized as a hybrid index (HI) and quantifies genome-wide variation in introgression among putatively admixed populations. The model includes two cline parameters that determine the probability that an individual with a given HI inherited a gene copy at a locus from one parental population, ϕ, while the probability from the other parental population ancestry is $1- \phi$ (Gompert & Buerkle 2011b). The genomic cline parameter α gives the

cline center, which describes the increase or decrease in the probability of ancestry for one of the species at a particular locus given the hybrid index while the genomic cline parameter β describes the rate of transition from low to high probability of ancestry as a function of the hybrid index, indicating a steeper or shallower change in allele frequencies at a given locus (Gompert & Buerkle 2011b).

1.3.3. Evolutionary consequences of hybridization

The analysis of hybrid zones has demonstrated that distinct taxa can persist even in the face of gene flow, challenging the strict interpretations of the biological species concept (Harrison 1990; Mallet 2005; Harrison & Larson 2014). More than a consequence when two diverging populations interbreed, hybridization is an important evolutionary phenomenon.

Introgressive hybridization has a variable importance in the evolutionary history of species. Gene flow between diverging populations may culminate in the genome homogenization of both populations or rather the reproductive barriers can be reinforced (Arnold 1992). When in contact, the evolutionary trajectory of divergent populations depends on the genomic architecture of differentiation, i.e. number of loci involved, how strong is the adaptation of loci combinations in each genomic background, whether such combinations are involved in epistatic interactions, etc., and the amount of gene flow between divergent populations (Wu 2001).

Interspecific hybridization is typically strongly deleterious or may drive populations in the direction of decreasing differentiation. In taxa without strong postmating barriers to gene exchange experiencing interspecific hybridization, genetic admixture may lead to the loss of genetic variability and ultimately reverse speciation (Seehausen *et al.* 2008). For instance, under changes in ecological conditions, convergence in morphological traits and decrease in genetic differentiation was documented between two Darwin's finches *Geospiza* species and attributed to introgression and selection (Grant *et al.* 2004).

Nevertheless, it is currently well-known that invasions of the genome via introgressive hybridization can contribute to adaptability and diversification, allowing novel adaptive combinations to evolve at a higher rate than in the absence of an external input of genetic variation (Mallet 2005). Around 2 to 4% of cases of speciation in

flowering plants and 7% in fern species are linked with polyploidization (events originating more than two homologous chromosome copies), from which many of these cases are thought to be by allopolyploidization, i.e. chromosome duplications in F1 hybrids (Otto & Whitton 2000). Despite being less common, speciation driven by hybridization has also been shown to occur in animals (Orr 1990; Otto & Whitton 2000). There are cases in which introgressive hybridization added new external adaptive traits. Melanism in North American populations of grey wolf *Canis lupus* was caused by past hybridization with domestic dog showing sign of positive selection in forest habitats (Anderson *et al.* 2009). Furthermore, hybridization is suggested to have influenced adaptive radiation across Darwin's finches from genus *Geospiza* by increasing variability of adaptive traits within species (Freeland & Boag 1999).

Another phenomenon that may contribute to an extensive input of genetic variation, whether beneficial or neutral, is cytoplasmic gene flow. Extensive mitochondrial and chloroplast introgression in nature are well-known consequences of introgressive hybridization (Senjo *et al.* 1999; Palme *et al.* 2004; Renoult *et al.* 2009; Melo-Ferreira *et al.* 2011). However, unfavourable effects of cytoplasmic gene flow are also known, for example, cytonuclear incompatibilities causing a decrease in hybrid fitness (Pereira *et al.* 2014) or sterility (Chase 2007).

Introgressive hybridization can thus influence the direction of evolution by instigating homogenization of divergent populations or by causing a genetic swamping of one by another, by the reinforcement of reproductive isolation, throughout the transference of genetic material between populations, or even originating new species. However, introgression may not have an obvious impact in the overall genetic composition of divergent populations if recombinant genotypes remain restricted to a narrow region as it happens in tension zones (Barton & Hewitt 1985). Hybrid genotypes remain trapped in restrict areas if they are less fit in parental habitats (Burke & Arnold 2001), due to endogenous or exogenous factors (Barton 2001).

1.4. Next-generation Sequencing data for the study of hybrid zones

Next-generation sequencing (NGS) technologies have revolutionized several research fields, including evolutionary biology. It is now possible to perform genome-wide studies

in organisms for which few genomic resources currently exist, which allowed to identify several thousands of informative SNPs in non-model organisms quickly and at reasonable costs. This kind of data can be generated for instance with a genotyping-by-sequencing (GBS; Elshire *et al.* 2011), restriction site associated DNA sequencing (RADseq; Baird *et al.* 2008; Hohenlohe *et al.* 2010, 2011) or similar methods without a prior knowledge on the genome of the studied organisms. SNPs are the most abundant type of genetic marker across the genome and their high density makes them ideal for studying hybridization. The increased numbers of SNP markers produced by genomic techniques and new analytical methods have been proved to be useful for the study of hybrid zones, not only for the estimation of hybrid indexes, individual admixture proportions and interspecific ancestry (Hohenlohe *et al.* 2011; Pujolar *et al.* 2014) but also to measure the variation of introgression across the genome (Parchman *et al.* 2013; Mandeville *et al.* 2015). The latter is useful in the identification of regions of the genome under selection and hence potentially involved in reproductive isolation (Harrison & Larson 2016; Gompert *et al.* 2017). Furthermore, a reference genome is particularly valuable to identify the genomic regions potentially in reproductive isolation, identify blocks of distinct ancestry in hybrids and regions with distinct levels of introgression (Nadeau *et al.* 2014; Maroja *et al.* 2015).

RAD sequencing offers attractive features to obtain a relatively large number of markers, widely and randomly distributed across genome. RAD tags create a reduced representation of the genome, allowing oversequencing of the nucleotides next to restriction sites and detection of a suitable number of SNPs by the choice of a restriction enzyme (Baird *et al.* 2008; Davey *et al.* 2011; Hohenlohe *et al.* 2011). By using additional enzymes, the number of RAD tags may be increased (Baird *et al.* 2008; Davey *et al.* 2011; Hohenlohe *et al.* 2011). RAD sequencing through the use of Illumina® next-generation sequencing allows to simultaneously discover and score thousands to hundreds of thousands of SNPs for a relatively reduced cost while parallel and multiplexed sample sequencing facilitates the rapid and high-throughput genotyping of large number of individuals (Baird *et al.* 2008; Davey *et al.* 2011; Hohenlohe *et al.* 2011).

1.5. The study model: *Podarcis hispanicus* complex

Wall lizards of the genus *Podarcis* are naturally circumscribed to the Mediterranean

Basin (Arnold *et al.* 2007), occupying a wide variety of habitats from Central Europe to the northern limits of the Sahara and from the Iberian Peninsula to the Black Sea (Sillero et al., 2014). Species of *Podarcis* from the Iberian Peninsula and Northern Africa, with the exception of *Podarcis muralis* (Laurenti, 1768), form a monophyletic clade (Harris & Arnold 1999; Oliverio *et al.* 2000), which constitutes the *Podarcis hispanicus* (Steindachner, 1870) complex (Harris & Sá-Sousa 2002). To date, several phylogenetic studies on the *P. hispanicus* group (e.g. Harris & Sá-Sousa 2002; Pinho *et al.* 2006; Lima *et al.* 2009; Kaliontzopoulou *et al.* 2011) have revealed cryptic variation and led to the discovery of 16 mitochondrial (mtDNA) lineages, including 11 lineages in the Iberian Peninsula and five in North Africa with a wide range of divergence times (TMRCA ranging from 2.69 to 10.41 Mya; Kaliontzopoulou *et al.* 2011). Eight lineages, identified on the basis of multilocus genetic data (see Pinho *et al.* 2007), morphology and ecology, are currently recognized as valid species or subspecies (see Pérez-Mellado 1981; Sá-Sousa 2001; Sá-Sousa & Harris 2002; Busack *et al.* 2005; Geniez *et al.* 2007, 2014; Renoult *et al.* 2009).

Lizards from the *P. hispanicus* complex are widely distributed in the Iberomaghrebian region, occupying a high diversity of ecological conditions. Among this group most pairs have predominantly allopatric distributions but some pairs may be found in parapatry or even sympatry throughout these regions. Besides sympatric forms, parapatric forms occasionally come into contact in restricted contact zones.

Robust phylogenies (Lima *et al.* 2009; Kaliontzopoulou *et al.* 2011), genotype assignment (Pinho *et al.* 2007) and morphological analysis (Kaliontzopoulou *et al.* 2012) support the existence of such highly differentiated evolutionary units mentioned above but discordance among loci caused by incomplete lineage sorting and/or gene flow emphasizes the incomplete divergence between some forms and the gradual evolution of reproductive isolation (Pinho *et al.* 2007, 2008). Several *Podarcis* forms seem to be prone to natural hybridization with congenerics (Capula 1993, 2002). A study of the only contact zone between *P. bocagei* and *P. carbonelli* showed that the two species occasionally hybridize but also indicated the existence of strong reproductive isolation, as a large fraction of individuals in the populations of contact was assigned to one of the two species (Pinho *et al.* 2009), clearly showing that the hybrid zone is bimodal (Jiggins & Mallet 2000). One mitochondrial lineage (called “Valencia” by Renoult *et al.*, 2009 or “hispanicus sensu stricto” by Kaliontzopoulou *et al.* 2011) has so far only been identified

as an introgressed lineage in all southern populations of *P. liolepis* and some populations of *P. hispanicus* (Renoult *et al.* 2009). Such results put in evidence that the evolution of reproductive isolation, between at least some *P. hispanicus* forms, occurred or is still occurring in the presence of gene flow. Given the large range of differentiation and divergence times, together with the diversity of geographical situations, the *Podarcis hispanicus* complex is thus an interesting group to study evolutionary issues and test hypotheses related with speciation.

1.6. General objectives and thesis organization

In this thesis we first investigated the influence of ecological factors in shaping species distribution and thus its potencial influence for the establishment contact zones between all pairs of mtDNA lineages. Then we explored genomic tools to understand how reproductive isolation varies across a geographical gradient and across the genome in late stages of speciation. Finally multiple hybrid zones between distinct *Podarcis hispanicus* species were investigated to assess if gene flow is a common phenomenon between several pairs and if the degree of reproductive isolation is similar among interacting taxa. This knowledge will give us a comprehensive view on isolation mechanisms and the processes by which isolation is maintained in the late stages of speciation.

The structure of this thesis is divided in five chapters. Chapter I contains some background information about evolutionary processes prompting divergence between populations, leading to the evolution of speciation and the establishment of reproductive isolation. It also includes information concerning the limitations to species geographic ranges and the underlying causes, which is related with the establishment of contact zones between diverging populations. Subsequently hybrid zones are defined, and it is illustrated how these regions can be used to study speciation. Further information about current trends and methods on genomic analysis applied to the study of hybrid zones and an introduction to the model system are also shown, providing an outline for the studies that compose this thesis.

The next two chapters describe the studies on the two research fields followed in this work. Chapter II depicts the importance of niche divergence based on climatic and topographic factors and inter-specific competition within *P. hispanicus* complex for each

lineage distribution and co-occurrence. This chapter includes a first article entitled "*Lack of congruence of genetic and niche divergence in* Podarcis hispanicus *complex*", which has been published in the *Journal of Zoological Systematics and Evolutionary Research*. The main objective of this work was to assess if *Podarcis hispanicus* complex' lineages evolved in similar or distinct ecological conditions and the role of such conditions in the organization of current patterns of diversity across the distribution range. Ecological niche models were used to characterize the realized niche of each lineage based on topographic and climatic variables, to identify important variables, and to test for niche conservatism or divergence between pairs of lineages. Then phylogenetic relationships and niche similarity between pairs of lineages were compared to examine patterns of niche divergence. The results supported the hypothesis that genetic divergence across *Podarcis hispanicus* complex likely occurred in allopatric conditions, mostly with significant niche divergence. However, the common occurrence of partial niche overlap between lineages that exhibit strictly parapatric distributions suggests competition after secondary contact. The almost continuous distribution of *Podarcis* lizards in the study area appears to be a result of a combination of complementary suitable niches and competition, two important mechanisms contributing to the delimitation of geographic distributions and restricting the existence of extensive contact zones.

Chapter III is dedicated to the analysis of four of these contact zones including two more articles. One is entitled "*Genome-wide patterns of interspecific admixture in a natural hybrid zone between two* Podarcis *wall lizards in the late stages of speciation*" and is currently in preparation. In this study we analysed the only natural hybrid zone between *Podarcis bocagei* and *P. carbonelli*. Previous studies in this contact zone have shown that these two species form a narrow and bimodal hybrid zone but the low number of loci used did not allow a fine-scale examination of the genomic patterns of introgression. With this study we intended to understand how reproductive isolation varies across the genome and to identify regions potentially involved in reproductive isolation between these two species. A RADseq genotyping approach was used, resulting in a set of thousands of SNP markers to examine the extent of hybridization, level of admixture and variation in selection against introgression among loci. The results showed that interspecific gene flow is restricted to the narrow contact zone without extensive introgression into the parental populations. The genomic analysis showed that loci with restricted introgression are distributed across the genome, implying that

selection against gene flow occurs across the genome rather than being restricted to few genomic regions; a particular role for the Z chromosome in the reproductive isolation is also suggested.

The other article is entitled "*Evolution of sympatry without complete reproductive isolation: does hybridization matter for* Podarcis carbonelli *conservation?*" and is in preparation. Here we analysed four contact zones between *P. carbonelli* and four other *Podarcis* species. The main objectives were to understand if *P. carbonelli* could maintain gene flow with other co-occurring species, to compare patterns and investigate possible consequences of interspecific gene flow and to evaluate how to account hybridization for the conservation of this species, an endangered endemism from the Iberian Peninsula. The strong bimodality across most contact zones without obvious ecological and temporal isolation suggested strong intrinsic prezygotic isolation, like assortative mating. However, the recurrent gene flow between *P. carbonelli* and three other congeneric species, was confirmed despite deep divergence between them. *P. carbonelli* evolved without complete reproductive isolation between most species that co-occur. This species is listed as endangered and urgent conservation measures are needed, incorporating the assessment of the consequences of hybridization.

Chapter IV contains a general discussion exposing the major results and contextualizing the outcomes of this thesis in the study of speciation. First it is important to understand how ecological divergence shapes *Podarcis* distributions, by relating the degree of niche and genetic divergence. Niche divergence, or lack thereof, can contribute to establish barriers to gene flow or to the formation of contact zones, respectively, but interspecific competition is suggested to play a role in avoiding extensive range overlap. Then it is presented how reproductive isolation between two *Podarcis* species was built across the genome by assessing genomic patterns of hybridization and introgression. Extrinsic barriers, for example ecological conditions as those identified in the first study, seem to have an important role in reproductive isolation. Intrinsic barriers to interspecific gene flow are spread throughtout the genome and likely influenced by complex combinations of several forms of selection or even neutral processes. A possible distinct role in reprodutive isolation for the Z chromosome (compared to autosomes) is discussed. Finally a comparative study of four contact zones all involving *P. carbonelli* revealed that this species evolved without complete reproductive isolation between most of co-occuring species. Since this species is

sympatric throughtout its distribution, at least with one other species, and is considered as endangered, hybridization may have several implications and urgent conservation measures are needed, including the assessment of the consequences of hybridization in management actions.

The last chapter consists of the major conclusions achieved with this project and the future research directions.

1.7. References

Anderson TM, Candille SI, Musiani M *et al.* (2009) Molecular and evolutionary history of melanism in North American gray wolves. *Science*, **323**, 1339–1343.

Arnold ML (1992) Natural hybridization as an evolutionary process. *Annual review of Ecology and Systematics*, **23**, 237–261.

Avise JC (1998) Pleistocene phylogeographic effects on avian populations and the speciation process. *Proceedings of the Royal Society of London B: Biological Sciences*, **265**, 457–463.

Avise JC, Walker D, Johns GC (1998) Speciation durations and Pleistocene effects on vertebrate phylogeography. *Proceedings of the Royal Society of London B: Biological Sciences*, **265**, 1707–1712.

Baird SJE (1995) A simulation study of multilocus clines. *Evolution*, **49**, 1038–1045.

Baird NA, Etter PD, Atwood TS *et al.* (2008) Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PloS one*, **3**, e3376.

Baldassarre DT, White TA, Karubian J, Webster MS (2014) Genomic and morphological analysis of a semipermeable avian hybrid zone suggests asymmetrical introgression of a sexual signal. *Evolution*, **68**, 2644–2657.

Barbosa AM, Sillero N, Martínez-Freiría F, Real R (2012) Ecological niche models in Mediterranean herpetology: past, present, and future. *Ecological Modeling. Nova Science Publishers, Hauppauge, NY*, 173–204.

Barton NH (1979) The dynamics of hybrid zones. *Heredity*, **43**, 341.

Barton NH, Gale KS (1993) Genetic analysis of hybrid zones. *Hybrid zones and the evolutionary process*, 13–45.

Barton N, Hewitt G (1981) The genetic basis of hybrid inviability in the grasshopper *Podisma pedestris*. *Heredity*, **47**, 367–383.

Barton NH, Hewitt GM (1985) Analysis of hybrid zones. *Annual review of Ecology and Systematics*, **16**, 113–148.

Barton NH, Hewitt GM (1989) Adaptation, speciation and hybrid zones. *Nature*, **341**, 497.

Bateson W (1909) Heredity and variation in modern lights. Pp. 85-101, in: Seward AC (ed.), Darwin and Modern Science, Cambridge University Press.

Bazykin AD (1969) Hypothetical mechanism of speciaton. *Evolution*, **23**, 685–687.

Britch SC, Cain ML, Howard DJ (2001) Spatio-temporal dynamics of the *Allonemobius fasciatus*–A. socius mosaic hybrid zone: a 14-year perspective. *Molecular Ecology*, **10**, 627–638.

Bull CM, Burzacott D (2001) Temporal and spatial dynamics of a parapatric boundary between two Australian reptile ticks. *Molecular Ecology*, **10**, 639–648.

Butlin RK, Ritchie MG (2009) Genetics of speciation. *Heredity*, **102**, 1–3.

Chatfield MWH, Kozak KH, Fitzpatrick BM, Tucker PK (2010) Patterns of differential introgression in a salamander hybrid zone: Inferences from genetic data and ecological niche modelling. *Molecular Ecology*, **19**, 4265–4282.

Chase CD (2007) Cytoplasmic male sterility: a window to the world of plant mitochondrial–nuclear interactions. *TRENDS in Genetics*, ***23***, 81–90.

Coyne JA (1992) Genetics and speciation. *Nature*, **355**, 511.

Coyne JA, Orr HA (1989) Patterns of speciation in *Drosophila*. *Evolution*, **43**, 362–381.

Coyne JA, Orr HA (1998) The evolutionary genetics of speciation. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **353**, 287–305.

Dasmahapatra KK, Blum MJ, Aiello A *et al.* (2002) Inferences from a rapidly moving hybrid zone. *Evolution*, **56**, 741–753.

Dobzhansky TG (1937) *Genetics and the Origin of Species*. Columbia university press.

Elshire RJ, Glaubitz JC, Sun Q *et al.* (2011) A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PloS One*, **6**, e19379.

Endler JA (1977) *Geographic variation, speciation, and clines*. Princeton University Press.

Excoffier L, Dupanloup I, Huerta-Sánchez E, Sousa VC, Foll M (2013) Robust demographic inference from genomic and SNP data. *PLoS Genetics*, **9**.

Fitzpatrick BM (2013) Alternative forms for genomic clines. *Ecology and Evolution*, **3**, 1951–1966.

Fitzpatrick BM, Johnson JR, Kump DK *et al.* (2009) Rapid fixation of non-native alleles revealed by genome-wide SNP analysis of hybrid tiger salamanders. *BMC Evolutionary Biology*, **9**, 176.

Freeland JR, Boag PT (1999) The mitochondrial and nuclear genetic homogeneity of the phenotypically diverse Darwin's ground finches. *Evolution*, **53**, 1553–1563.

Gapare WJ, Aitken SN, Ritland CE (2005) Genetic diversity of core and peripheral Sitka spruce (*Picea sitchensis* (Bong.) Carr) populations: implications for conservation of widespread species. *Biological Conservation*, **123**, 113–123.

Garner TWJ, Pearman PB, Angelone S (2004) Genetic diversity across a vertebrate species' range: a test of the central–peripheral hypothesis. *Molecular Ecology*, **13**, 1047–1053.

Gaston KJ (2009) Geographic range limits: achieving synthesis. *Proceedings of the Royal Society of London B: Biological Sciences*, rspb-2008.

Gay L, Crochet PA, Bell DA, Lenormand T (2008) Comparing clines on molecular and phenotypic traits in hybrid zones: A window on tension zone models. *Evolution*, **62**, 2789–2806.

Gómez A, Lunt DH (2007) Refugia within refugia: patterns of phylogeographic concordance in the Iberian Peninsula. In: *Phylogeography of southern European refugia*, pp. 155–188. Springer.

Gompert Z, Buerkle CA (2009) A powerful regression-based method for admixture mapping of isolation across the genome of hybrids. *Molecular Ecology*, **18**, 1207–1224.

Gompert Z, Buerkle CA (2011a) Bayesian estimation of genomic clines. *Molecular Ecology*, **20**, 2111–2127.

Gompert Z, Buerkle CA (2011b) A hierarchical bayesian model for next-generation population genomics. *Genetics*, **187**, 903–917.

Gompert Z, Lucas LK, Buerkle CA *et al.* (2014) Admixture and the organization of genetic diversity in a butterfly species complex revealed through common and rare genetic variants. *Molecular Ecology*, **23**, 4555–4573.

Gompert Z, Mandeville EG, Buerkle CA (2017) Analysis of population genomic data from hybrid zones. *Annual Review of Ecology, Evolution, and Systematics*, **48**, 207–229.

Gompert Z, Parchman TL, Buerkle CA (2012) Genomics of isolation in hybrids.

Philosophical Transactions of the Royal Society B: Biological Sciences, **367**, 439–450.

Grant PR, Grant BR, Markert JA, Keller LF, Petren K (2004) Convergent evolution of Darwin's finches caused by introgressive hybridization and selection. *Evolution*, **58**, 1588–1599.

Gravel S (2012) Population genetics models of local ancestry. *Genetics*, **191**, 607–619.

Guisan A, Thuiller W (2005) Predicting species distribution: Offering more than simple habitat models. *Ecology Letters*, **8**, 993–1009.

Harrison RG (1990) Hybrid zones: Windows on evolutionary process. *Oxford surveys in evolutionary biology*, **7**, 69–128.

Harrison RG (1993) Hybrids and hybrid zones: historical perspective. *Hybrid zones and the evolutionary process*, 3–12.

Harrison RG, Larson EL (2014) Hybridization, introgression, and the nature of species boundaries. *Journal of Heredity*, **105**, 795–809.

Harrison RG, Larson EL (2016) Heterogeneous genome divergence, differential introgression, and the origin and structure of hybrid zones. *Molecular Ecology*, **25**, 2454–2466.

Hewitt GM (1988) Hybrid zones - natural laboratories for evolutionary studies. *Trends in Ecology & Evolution*, **3**, 158–167.

Hewitt GM (2001) Speciation, hybrid zones and phylogeography - or seeing genes in space and time. *Molecular Ecology*, **10**, 537–549.

Hey J (2001) The mind of the species problem. *Trends in Ecology & Evolution*, **16**, 326–329.

Hey J (2006a) On the failure of modern species concepts. *Trends in Ecology & Evolution*, **21**, 447–450.

Hey J (2006b) Recent advances in assessing gene flow between diverging populations and species. *Current Opinion in Genetics & Development*, **16**, 592–596.

Hey J, Won Y, Sivasundar A, Nielsen R, Markert JA (2004) Using nuclear haplotypes with microsatellites to study gene flow between recently separated Cichlid species. *Molecular Ecology*, **13**, 909–919.

Hohenlohe PA, Amish SJ, Catchen JM, Allendorf FW, Luikart G (2011) Next-generation RAD sequencing identifies thousands of SNPs for assessing hybridization between rainbow and westslope cutthroat trout. *Molecular Ecology Resources*, **11**, 117–

122.

Hohenlohe PA, Bassham S, Etter PD *et al.* (2010) Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. *PLoS Genetics*, **6**.

Kambhampati S, Rai KS, Burgun SJ (1993) Unidirectional cytoplasmic incompatibility in the mosquito, *Aedes albopictus*. *Evolution*, **47**, 673–677.

Key KHL (1968) The concept of stasipatric speciation. *Systematic Biology*, **17**, 14–22.

Kondrashov AS (1983a) Multilocus model of sympatric speciation I. One character. *Theoretical Population Biology*, **24**, 121–135.

Kondrashov AS (1983b) Multilocus model of sympatric speciation II. Two characters. *Theoretical Population Biology*, **24**, 136–144.

Lexer C, Buerkle CA, Joseph JA, Heinze B, Fay MF (2007) Admixture in European *Populus* hybrid zones makes feasible the mapping of loci that contribute to reproductive isolation and trait differences. *Heredity*, **98**, 74.

Mallet J (2005) Hybridization as an invasion of the genome. *Trends in Ecology and Evolution*, **20**, 229–237.

Mallet J (2008) Hybridization, ecological races and the nature of species: empirical evidence for the ease of speciation. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **363**, 2971–2986.

Mallet J, Barton N, Lamas G *et al.* (1990) Estimates of selection and gene flow from measures of cline width and linkage disequilibrium in *Heliconius* hybrid zones. *Genetics*, **124**, 921–936.

Mandeville EG, Parchman TL, McDonald DB, Buerkle CA (2015) Highly variable reproductive isolation among pairs of *Catostomus* species. *Molecular Ecology*, **24**, 1856–1872.

Maroja LS, Larson EL, Bogdanowicz SM, Harrison RG (2015) Genes with restricted introgression in a field cricket (*Gryllus firmus*/*Gryllus pennsylvanicus*) hybrid zone are concentrated on the X chromosome and a single autosome. *G3: Genes, Genomes, Genetics*, **5**, 2219–2227.

Marshall JL (2004) The *Allonemobius*-*Wolbachia* host-endosymbiont system: evidence for rapid speciation and against reproductive isolation driven by cytoplasmic incompatibility. *Evolution*, **58**, 2409–2425.

Mayden RL (1997) A hierarchy of species concepts: the denouement in the saga of the

species problem.

Mayr E (1942) *Systematics and the origin of species, from the viewpoint of a zoologist*. Harvard University Press.

Mayr E (1963) *Animal species and evolution*. Belknap Press. Cambridge, Massacusetts.

Melo-Ferreira J, Boursot P, Carneiro M *et al.* (2011) Recurrent introgression of mitochondrial DNA among hares (*Lepus* spp.) revealed by species-tree inference and coalescent simulations. *Systematic Biology*, **61**, 367–381.

Le Moan A, Gagnaire P, Bonhomme F (2016) Parallel genetic divergence among coastal–marine ecotype pairs of European anchovy explained by differential introgression after secondary contact. *Molecular Ecology*, **25**, 3187–3202.

Muller H (1942) Isolating mechanisms, evolution, and temperature. In: *Biol. Symp.*, pp. 71–125.

Nadeau NJ, Ruiz M, Salazar P *et al.* (2014) Population genomics of parallel hybrid zones in the mimetic butterflies, *H. melpomene* and *H. erato*. *Genome Research*, **24**, 1316–1333.

Navarro A, Barton NH (2003) Chromosomal speciation and molecular divergence-accelerated evolution in rearranged chromosomes. *Science*, **300**, 321–324.

Nosil P (2008) Speciation with gene flow could be common. *Molecular Ecology*, **17**, 2103–2106.

Nosil P, Schluter D (2011) The genes underlying the process of speciation. *Trends in Ecology & Evolution*, **26**, 160–167.

Orr HA (1990) “Why polyploidy is rarer in animals than in plants” revisited. *The American Naturalist*, **136**, 759–770.

Otto SP, Whitton J (2000) Polyploid incidence and evolution. *Annual Review of Genetics*, **34**, 401–437.

Palme AE, Su Q, Palsson S, Lascoux M (2004) Extensive sharing of chloroplast haplotypes among European birches indicates hybridization among *Betula pendula*, *B. pubescens* and *B. nana*. *Molecular Ecology*, **13**, 167–178.

Parchman TL, Gompert Z, Braun MJ *et al.* (2013) The genomic consequences of adaptive divergence and reproductive isolation between species of manakins. *Molecular Ecology*, **22**, 3304–3317.

Payseur BA (2010) Using differential introgression in hybrid zones to identify genomic regions involved in speciation. *Molecular Ecology Resources*, **10**, 806–820.

Payseur BA, Krenz JG, Nachman MW (2004) Differential patterns of introgression across the X chromosome in a hybrid zone between two species of house mice. *Evolution*, **58**, 2064–2078.

Pereira RJ, Barreto FS and Burton RS (2014) Ecological novelty by hybridization: experimental evidence for increased thermal tolerance by transgressive segregation in *Tigriopus californicus*. *Evolution*, ***68***, 204–215.

Pujolar JM, Jacobsen MW, Als TD *et al.* (2014) Assessing patterns of hybridization between North Atlantic eels using diagnostic single-nucleotide polymorphisms. *Heredity*, **112**, 627.

Rand DM, Harrison RG (1989) Ecological genetics of a mosaic hybrid zone: mitochondrial, nuclear, and reproductive differentiation of crickets by soil type. *Evolution*, **43**, 432–449.

Renoult JP, Geniez P, Bacquet P, Benoit L, Crochet PA (2009) Morphology and nuclear markers reveal extensive mitochondrial introgressions in the Iberian Wall Lizard species complex. *Molecular Ecology*, **18**, 4298–4315.

Rieseberg LH (2001) Chromosomal rearrangements and speciation. *Trends in Ecology & Evolution*, **16**, 351–358.

Rieseberg LH, Blackman BK (2010) Speciation genes in plants. *Annals of Botany*, **106**, 439–455.

Rundle HD, Nosil P (2005) Ecological speciation. *Ecology Letters*, **8**, 336–352.

Saarman NP, Pogson GH (2015) Introgression between invasive and native blue mussels (genus *Mytilus*) in the central California hybrid zone. *Molecular Ecology*, **24**, 4723–4738.

Sedghifar A, Brandvain Y, Ralph P (2016) Beyond clines: lineages and haplotype blocks in hybrid zones. *Molecular Ecology*, **25**, 2559–2576.

Seehausen OLE, Takimoto G, Roy D, Jokela J (2008) Speciation reversal and biodiversity dynamics with hybridization in changing environments. *Molecular Ecology*, **17**, 30–44.

Senjo M, Kimura K, Watano Y, Ueda K, Shimizu T (1999) Extensive mitochondrial introgression from *Pinus pumila* to *P. parviflora* var. pentaphylla (Pinaceae). *Journal of Plant Research*, **112**, 97–105.

Sequeira F, Alexandrino J, Rocha S, Arntzen JW, Ferrand N (2005) Genetic exchange across a hybrid zone within the Iberian endemic golden-striped salamander,

Chioglossa lusitanica. *Molecular Ecology*, **14**, 245–254.

Slatkin M (1973) Gene flow and selection in a cline. *Genetics*, **75**, 733–756.

Soberón J (2007) Grinnellian and Eltonian niches and geographic distributions of species. *Ecology Letters*, **10**, 1115–1123.

Soberón J, Peterson AT (2005) Interpretation of models of fundamental ecological niches and species' distributional areas. *Biodiversity Informatics*, **2**, 1–10.

Szymura JM, Barton NH (1986) Genetic analysis of a hybrid zone between the fire-bellied toads, *Bombina bombina* and *B. variegata*, near Cracow in southern Poland. *Evolution*, **40**, 1141–1159.

Szymura JM, Barton NH (1991) The genetic structure of the hybrid zone between the fire-bellied toads *Bombina bombina* and *B. variegata*: comparisons between transects and between loci. *Evolution*, **45**, 237–261.

Tarroso P, Pereira RJ, Martínez-Freiría F, Godinho R, Brito JC (2014) Hybridization at an ecotone: Ecological and genetic barriers between three Iberian vipers. *Molecular Ecology*, **23**, 1108–1123.

Taylor SA, White TA, Hochachka WM *et al.* (2014) Climate-mediated movement of an avian hybrid zone. *Current Biology*, **24**, 671–676.

Telschow A, Flor M, Kobayashi Y, Hammerstein P, Werren JH (2007) *Wolbachia*-induced unidirectional cytoplasmic incompatibility and speciation: mainland-island model. *PLoS One*, **2**, e701.

Via S (2001) Sympatric speciation in animals: the ugly duckling grows up. *Trends in Ecology & Evolution*, **16**, 381–390.

While G M, Michaelides S, Heathcote RJ, MacGregor HE, Zajac, N., Beninde J, *et al.* (2015). Sexual selection drives asymmetric introgression in wall lizards. *Ecology Letters*, *18*, 1366–1375.

Wielstra B, Burke T, Butlin RK *et al.* (2017) A genomic footprint of hybrid zone movement in crested newts. *Evolution Letters*, **1**, 93–101.

Wilding CS, Butlin RK, Grahame J (2001) Differential gene exchange between parapatric morphs of *Littorina saxatilis* detected using AFLP markers. *Journal of Evolutionary Biology*, **14**, 611–619.

Wu CI (2001) The genic view of the process of speciation. *Journal of Evolutionary Biology*, **14**, 851–865.

Wu CI, Ting CT (2004) Genes and speciation. *Nature Reviews Genetics*, **5**, 114.

Zbawicka M, Sańko T, Strand J, Wenne R (2014) New SNP markers reveal largely concordant clinal variation across the hybrid zone between *Mytilus* spp. in the Baltic Sea. *Aquatic Biology*, **21**, 25–36.

Chapter 2. Importance of climatic niche divergence and inter-specific competition for species distribution and co-occurrence within *P. hispanicus* complex

Article I. Lack of congruence of genetic and niche divergence in *Podarcis hispanicus* complex.

Caeiro-Dias G, Luís C, Pinho C, Crochet PA, Sillero N, Kaliontzopoulou A (2018) *Journal of Zoological Systematics and Evolutionary Research*, **56**, 479-492. DOI: https://doi.org/10.1111/jzs.12219

Article I. Lack of congruence of genetic and niche divergence in *Podarcis hispanicus* complex

Guilherme Caeiro-Dias[1,2,3], Carla Luís[1], Catarina Pinho[1], Pierre-André Crochet[3], Neftalí Sillero[4*], Antigoni Kaliontzopoulou[1*]

[1]Centro de Investigação em Biodiversidade e Recursos Genéticos, CIBIO/InBio, Vairão, Portugal
[2]Departamento de Biologia, Faculdade de Ciências, Universidade do Porto, Porto, Portugal
[3]CEFE, CNRS, University of Montpellier, University Paul Valéry Montpellier 3, EPHE, IRD, Montpellier, France
[4]Centro de Investigação em Ciências Geo-Espaciais, CICGE, Observatório Astronómico Prof. Manuel de Barros, Vila Nova de Gaia, Portugal
*These authors contributed equally to this study.

Abstract

Niche divergence among closely related lineages can be informative on the ecological and evolutionary processes involved in differentiation, particularly in the case of cryptic species complexes. Here we compared phylogenetic relationships and niche similarity between pairs of lineages included in the *Podarcis hispanicus* complex to examine patterns of niche divergence and its role in the organization of current diversity patterns, as allopatric, parapatric, and sympatric lineages occur in the Western Mediterranean Basin. First, we used ecological niche models to characterize the realized climatic niche of each *Podarcis* hispanicus complex lineage based on topographic and climatic variables, to identify important variables, and to test for niche conservatism or divergence between pairs of lineages. Variables related to precipitation generally exhibited the highest contribution to niche models, highlighting the importance of rainfall levels in shaping distributions of *Podarcis* wall lizards. We found that most forms have significant differences in realized climatic niches that do not follow the pattern of mitochondrial divergence. These results lend support to the hypothesis that genetic divergence across *Podarcis hispanicus* complex most likely occurred in allopatric conditions, mostly with significant niche divergence. Competition after secondary contact is also suggested by the common occurrence of niche overlap between lineages that exhibit strictly parapatric distribution. The almost continuous distribution of *Podarcis*

lizards in the study area appears to be a result of a combination of complementary suitable niches and competition, which seem two important mechanisms limiting geographic distributions and restricting the existence of extensive contact zones.

Keywords: allopatric speciation; ecological niche modeling; *Lacertidae*; Maxent; mtDNA lineages; niche divergence.

Introduction

The geographic distribution of organisms is the result of three main limiting factors: determinants to dispersion (species' intrinsic or historical constraints), abiotic factors, and biotic interactions: Species can live in climatically favorable regions where they are able to disperse and from where they are not excluded by biotic interactions (Barbosa *et al.*, 2012; Soberón, 2007). Range limits are thus determined by the areas where ecological conditions are favorable: The species' range is the geographic expression of the species' ecological niche. Grinnell (1917) defined the ecological niche as a portion of the habitat containing the environmental conditions that enable individuals of a species to survive and reproduce based on broad-scale variables (climate) that are not affected by species density. Instead, Elton (1927) emphasized the functional role of a species in a community, especially its position in food webs, based on fine-scale variables representing resources that may be consumed or modified by the species. The fundamental niche is the multidimensional environmental space where each dimension is a variable describing the conditions under which a species can maintain a viable population and persist over time (Hutchinson, 1957). Jackson and Overpeck (2000) defined the concept of potential niche as the part of the fundamental niche that is currently available for the species and Pearson (2007) introduced the concept of occupied niche, to which species distributions are limited by historical, geographic, and biotic factors. The realized niche is the part of the fundamental niche where the species is not excluded by biotic interactions and dispersal factors, and it may inform us about environmental variables driving range limits (Holt, 2003; Soberón & Nakamura, 2009). Because of this multidimensionality, it is not possible to investigate into detail each and every one of the dimensions that influence the niche, but climatic and physical factors

merit particular attention, as they are recognized to affect profoundly the distributions of species at small scales (Soberón & Peterson, 2005).

We here follow Sillero (2011), who proposed the term realized niche model when predicting the species' realized niche, for those correlative models using presence/true-absence, presence/pseudoabsence, or presence-only species records. When using presenceonly data, like the analyses in this study, the model represents the realized niche and not the fundamental one, because historical/dispersal factors are in fact included in the presence data. Here we use mainly climatic variables associated with presence data, so we effectively model the realized climatic niche, while the projected distributions are the geographic extent of the realized climatic niche.

In view of the importance of the realized niche for understanding the environmental conditions that favor the occurrence of a species, comparing the realized climatic niche between pairs of species may also provide insights about other factors limiting distributions, as well as about the location and extent of contact zones (Rieseberg *et al.*, 2003; Swenson, 2006; Swenson *et al.*, 2008). When two species have widely overlapping realized climatic niches, but exhibit parapatric or allopatric distributions in the geographic space, we can infer that their distributions are limited by other factors (e.g., biotic, historical, dispersal: Warren *et al.* 2014), under the condition that the studied variables are important for describing the species' niche. Movement abilities and dispersal limitations play an important role in this topic (Holt, 2003). Considering the aforementioned parapatric example, biotic factors like a competition between the two species may also contribute to explaining distribution limits (Barbosa *et al.*, 2012; Soberón & Peterson, 2005). Similarly, reduced dispersal capabilities and the existence of geographic barriers may favor or facilitate the allopatric occurrence of two species (Barbosa *et al.*, 2012; Holt, 2003; Soberón & Peterson, 2005).

The role of ecological divergence in driving the origin and maintenance of distinct lineages has been the subject of long-standing interest in evolutionary biology (Wright, 1921; Mayr, 1963; Coyne and Orr, 2004). Divergence in ecological niche has been hypothesized to constitute one of the main mechanisms of speciation (Schluter, 2001, 2009). Under this hypothesis, adaptation to different environments is a major agent of the evolution of reproductive isolation. Divergent selection in ecology may occur in sympatry, as well as in a situation of allopatric or parapatric speciation. If the realized climatic niches of allopatric or parapatric sister lineages do not overlap, then the

ecological adaptation with barriers to dispersal may be inferred to have played an important role in lineage differentiation (Graham *et al.*, 2004; Wiens, 2004; Wiens & Graham, 2005). Alternatively, if realized climatic niches overlap significantly, speciation was not accompanied by divergence in environmental preferences and probably occurred in allopatry or parapatry (incidental divergence, e.g., genetic drift, Peterson, Soberón, & Sanchez-Cordero, 1999; Graham et al., 2004; Wiens, 2004; Wiens & Graham, 2005). In cases of non-sister lineages occupying similar niches, these may be inferred to have retained the ancestral niche through time or, alternatively, they may have independently derived the same niche through convergent evolution (Knouft *et al.*, 2006). Understanding the evolutionary patterns of niche diversification can thus disclose valuable insights about the distribution of lineages and its role in lineage divergence.

Wall lizards of the genus *Podarcis* evolved and diversified in the Mediterranean Basin (Arnold *et al.*, 2007), occupying a wide variety of habitats. They are distributed from Central Europe to the northern limits of the Sahara and from the Iberian Peninsula to the Black Sea (Sillero *et al.*, 2014). Species of *Podarcis* from the Iberian Peninsula and Northern Africa, with the exception of *Podarcis muralis* (Laurenti, 1768), form a monophyletic clade (Harris & Arnold, 1999; Oliverio *et al.*, 2000), which constitutes the *Podarcis hispanicus* (Steindachner, 1870) complex (Harris & Sá-Sousa, 2002). They are widely distributed in the Iberian Peninsula and often abundant in Northern Africa, where they exhibit high diversity of ecological situations, and several cases of sympatric, parapatric and allopatric distributions occur throughout these regions. The *Podarcis* hispanicus complex is thus an interesting group to study patterns of niche divergence and configuration of species' distributions.

The current range of Iberian and North African *Podarcis* lizards is mostly parapatric, with several restricted contact zones between parapatric lineages currently being investigated (e.g., *Podarcis bocagei* (Seoane, 1884) with *Podarcis carbonelli* Pérez-Mellado, 1981, Pinho *et al.*, 2009; *P. carbonelli* with *Podarcis vaucheri* (Boulenger, 1905), *Podarcis guadarramae guadarramae* (Boscá, 1916) with *Podarcis liolepis* (Boulenger, 1905), pers. obs.). Two lineages are mostly sympatric with other lineages (*P. bocagei* with *Podarcis guadarramae lusitanicus* (Geniez *et al.*, 2014), and *P. carbonelli* with *P. g. guadarramae*, *Podarcis virescens* (Geniez *et al.*, 2014), or *P. vaucheri*, Kaliontzopoulou *et al.*, 2011). Despite the increasing number of genetic studies on this complex of species, the ecological niche and distribution limits of some lineages

are still poorly known. During the last decade, the *Podarcis hispanicus* species complex has been subject to several phylogenetic studies (e.g., Harris & Sá-Sousa, 2002; Kaliontzopoulou *et al.*, 2011; Lima *et al.*, 2009; Pinho, *et al.*, 2006) which have uncovered cryptic variation and lead to the discovery of 16 mitochondrial lineages, including eleven lineages in the Iberian Peninsula and five in North Africa (Kaliontzopoulou *et al.*, 2011; Figure 2). Eight lineages, identified on the basis of multilocus genetic data (see Pinho *et al*., 2007), morphology and ecology, are currently recognized as valid species or subspecies (see Pérez-Mellado, 1981; Sá-Sousa, 2001; Sá-Sousa & Harris, 2002; Busack et al., 2005; Geniez, *et al*., 2007; Renoult, *et al*., 2009; Geniez *et al*., 2014), while others still lack taxonomic revision. Lastly, one mitochondrial (mtDNA) lineage (called "Valencia" by Renoult *et al*., 2009 or "*hispanicus sensu stricto*" by others) has so far only been identified as an introgressed lineage in all southern populations of liolepis and some population of *P. hispanicus* (Renoult *et al*., 2009). In this work, when we mention *P. liolepis* or *P. hispanicus* Galera, we refer only to Liolepis and Hispanicus mtDNA lineages mentioned by Renoult *et al*. (2009), respectively.

The application of ecological niche modeling (ENM) together with Geographical Information Systems (GIS) allows the development of robust models that relate biological diversity with environmental factors in a geographically explicit framework (Guisan & Zimmermann, 2000; Sillero, 2011; Warren, 2012, 2013). The integration of phylogenetic information allows assessing if closely related lineages evolved in similar or distinct ecological conditions (e.g., Kidd & Ritchie, 2006; Pearman *et al*., 2014; Rissler & Apodaca, 2007) either in sympatry, parapatry or allopatry. In this study, we employ ENM based on topographic and climatic variables, to (i) characterize the realized climatic niche (sensu Sillero, 2011) of each member of the *Podarcis hispanicus* species complex using genetically confirmed occurrence records and (ii) to identify which are the main topographic and climatic variables influencing the niche and how they are related to the presence of each lineage. Then, based on the inferred ecological niche models, we (iii) test for niche conservatism or divergence between lineages to assess the similarity (or lack thereof) in realized climatic niche and to incorporate these results with previously determined phylogenetic relationships within this group (see Kaliontzopoulou *et al*., 2011). Finally, we want to (iv) infer how niche divergence has contributed to current patterns of spatial organization among lineages of the *Podarcis hispanicus* complex and how it has shaped the potential for geographic co-occurrence between pairs of lineages.

Material and Methods

Study area and presence records

Our study area includes Southern France, the Iberian Peninsula, and the Mediterranean region of Northern Africa (Figure 2.1.), encompassing the known distribution range of the *Podarcis hispanicus* species complex. Due to the low number of available records for some lineages, we modeled together the monophyletic group formed by the three Algerian and Tunisian lineages (see Figure 2.2.), which we will refer to as Tunisia/Algeria group hereafter. The most recently discovered lineage of *P. vaucheri* from southern Spain (*P. vaucheri* SC Spain, Kaliontzopoulou *et al*., 2011) was excluded due to the very low number of geographic records. Because field morphological identification is still unsafe for several lineages, in this study we use mitochondrial lineages as a proxy for evolutionary lineages as this is the only cost-effective way to gather numerous genetically confirmed presence data. We thus assembled a total of 518 presence records for 13 *Podarcis* lineages identified on the basis of mtDNA. We have included 93 new samples specifically analyzed for this study (GenBank accession numbers KY461834 to KY461930) as well as 425 samples published in other studies, with mtDNA sequences and detailed geographic information available (Busack *et al*., 2005; Arntzen & Sá-Sousa, 2007; Carranza, Arnold, & Amat, 2004; Castilla et al., 1998; Dias *et al*., 2016; Harris & Arnold, 1999; Harris *et al*., 2002a; Harris *et al*., 2002b; Harris & Sá-Sousa, 2001, 2002; Kaliontzopoulou *et al*., 2011; Lima et al., 2009; Oliverio *et al*., 2000; Pinho, Harrid, & Ferrand, 2007; Pinho, Harris, & Ferrand, 2008; Pinho *et al*., 2006; Renoult, 2006; Sanz-Azkue, García-Etxebarria *et al*., 2006). For all individuals, we identified the mtDNA lineage through the analysis of partial fragments of mitochondrially encoded 12S ribosomal RNA region (12S), 16S ribosomal RNA region (16S), NADH dehydrogese subunit 4 (ND4), control region (D-loop), or cytochrome c oxidase subunit I (COI). Identification was performed by comparison of the target sequences with our extensive database of reference sequences available for all gene fragments employed using tree-building approaches like in Kaliontzopoulou *et al.* (2011). Individual identifications were all unambiguous given the good separation of all lineages in each of these gene fragments. For fragments specificaly sequenced for this study we used the primers and conditions described in Pinho, Ferrand & Harris (2006). The details on fragments sequenced for each sample are in Appendix I., Table S1.1. For several

individuals, the identification was based in more than one fragment (individuals with more than one fragment amplified are indicated with a "+" in Appendix I., Table S1.1.). Presence records were either geolocalized in the field with a GPS or the coordinates were retrieved from Google Earth for samples that were collected without GPS, with an accuracy according to the spatial resolution of environmental variables used for modeling purposes (see below). Detailed information about all used samples, as their respective assignments to mtDNA lineages and geographic coordinates, is listed in Appendix I., Table S1.1.

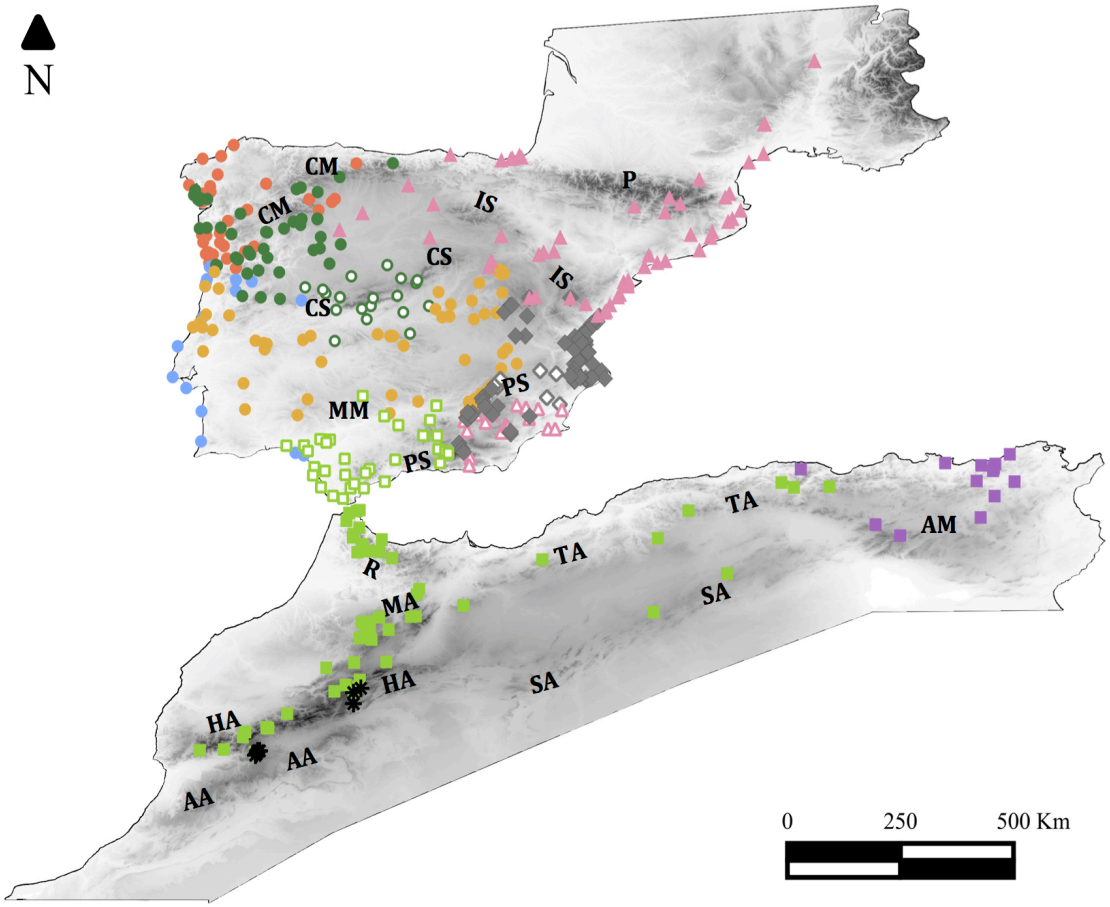

Figure 2.1. Presence records for each Podarcis hispanicus complex's lineage (red circles, P. bocagei (Pb); blue circles, P. carbonelli (Pc); green circles, P. guadarramae lusitanicus (Pgl); open circles, P. g. guadarramae (Pgg); orange circles, P. virescens (Pv); triangles, P. liolepis (Pl); open triangles, P. hispanicus Galera (PhG); gray diamonds, Valencia lineage (V); open diamonds, P. hispanicus Albacete/Murcia (PhAM); open squares, P. vaucheri Southern Spain (PvSS); light green squares, P. vaucheri Morocco/Algeria (PvMA); black asterisks, Jbel Siroua lineage (JS); purple squares, Tunisia/Algeria group (TAg)) and the location of the main mountain ranges in the study area (P, Pyrenees; CM, Cantabrian Mountains; IS, Iberian System; CS, Central System; MM, More Mountains; PS, Penibaetic System; R, Riff; MA; Middle Atlas; HA, High Atlas; AA, Anti Atlas; TA, Tell Atlas; SA, Saharan Atlas; AM, Aurés Mountains). Darker areas represent higher altitude.

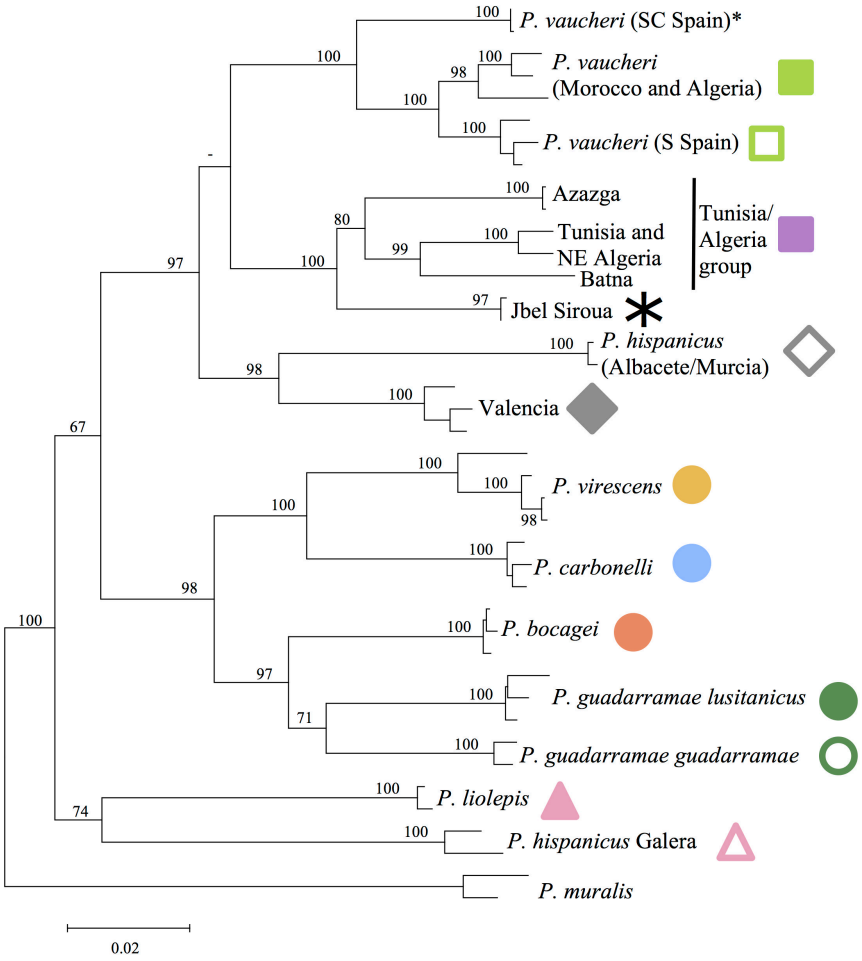

Figure 2.2. Maximum likelihood estimates of mtDNA relationships between Iberian and North African Podarcis. Podarcis muralis was used as outgroup. Lineage marked with * was not included in the analysis due to low number or samples. (Adapted from Kaliontzopoulou et al., 2011).

Environmental variables

A set of 21 variables was initially considered for the study area. Nineteen climatic variables were obtained from the WorldClim database (Hijmans et al., 2005; http://www.worldclim.org/). Two topographic variables were also obtained: Altitude was retrieved from the Shuttle Radar Topography Mission (USGS 2006; http://glcf.umd.edu/data/srtm/), and slope was calculated in QGIS version 2.4.0 (QGIS Development Team 2013). All layers had a spatial resolution of 1 km^2. To avoid the possible statistical effects of collinearity among predictor variables on niche modeling, we calculated pairwise Pearson correlations, and for each correlation higher than 0.75, we excluded the variable that correlates with the higher number of other variables, but always trying to keep the variables with a higher biological importance for the species. This resulted in a final set of seven variables considered for modeling purposes, including altitude (Alt), temperature seasonality (TSea), maximum temperature of the

warmest month (MaxT), minimum temperature of the coldest month (MinT), annual precipitation (APre), precipitation of the driest month (PreDM), and precipitation seasonality (PreSea). All spatial data operations and analyses were performed using QGIS version 2.4.0 (QGIS Development Team, 2013).

Niche modeling and characterization

To model the realized climatic niche (sensu Sillero, 2011) of each lineage, we used the maximum entropy approach implemented in Maxent version 3.4.1 (available from https://biodiversityinformatics.amnh.org/open_source/maxent/). Maxent was designed for presenceonly data (Phillips *et al*., 2006) and is particularly efficient for datasets including a small number of records (Pearson *et al*., 2007). We enabled linear, quadratic, and hinge features. Because of the small number of records available for some groups, we did not divide the datasets into training and testing subsets. As Maxent is a probabilistic method, each interation of the modeling process results in slightly different models. To account for model uncertainty, we calculated 50 replicated, independent Maxent models for each lineage, which were then averaged to obtain a consensus model (Araújo & New, 2007). For other parameters, we kept the default values. An important issue was whether our dataset was geographically biased or not. So we calculated models with and without correction for sampling bias. To take sampling bias into account when calculating Maxent models, we used a 100 km buffer area around each occurrence point where we randomly sampled 10,000 background samples as advocated by Phillips (2008) and Fourcade, Engler, Rödder, and Secondi (2014). We tested model performance by considering the area under the curve (AUC) of the receiver operating characteristics (ROC) plots (Liu *et al*., 2005). Variable importance was determined by jack-knife resampling of the model gain and AUC. Gain is a measure of how much better the predicted distribution fits the sample points as compared to a uniform distribution (Phillips & Dudík, 2008; Phillips *et al*., 2004; Phillips *et al*., 2006). For this purpose, Maxent excludes each variable in turn and creates a model with the remaining variables; then, it also creates a series of models using each variable separately. Maxent determines the contribution of each environmental variable to the final model by randomly permuting the values of each variable among the presence points and measuring the relative decrease in AUC, normalized to percentage (permutation importance).

In addition to Maxent models, we performed an ecological niche factor analysis (ENFA; Hirzel, *et al.*, 2002) to obtain a multivariate representation of the realized climatic niche of the different lineages by comparing the statistical distribution of climatic variables for each presence dataset (the lineage niche) and for the whole study area. This analysis was performed using Biomapper 4.0 (Hirzel, *et al*., 2006) following Hirzel *et al.* (2002). ENFA summarizes climatic variation into uncorrelated factors in a similar manner to Principal Component Analysis. The first factor, marginality, expresses the average difference between the species niche and the total available conditions (Hirzel *et al*., 2002). Marginality varies most often from 0, for species living in average habitat conditions, to 1, for species inhabiting very particular conditions relatively to the studied area. Larger values indicate that the mean variable values where species are present are fairly different from the average variable values across the studied area. The other factors are called specialization and represent the variance of the species niche compared with available conditions, ranging from 1 in generalist species, to infinite in specialist species (Hirzel *et al*., 2002). Comparisons between species are usually performed using the inverse of specialization, tolerance, which varies from 0 for specialist species to 1 for generalist species. The robustness of the models was evaluated with the continuous Boyce index using a Spearman rank correlation coefficient that measures how much model predictions differ from a random distribution of the observed presences across the prediction gradients (Boyce *et al*., 2002).

Niche differentiation

We analyzed climatic niche differentiation among lineages using the ENMTools 0.1 R package (Warren, 2016) in R 3.2.3 (R Development Core Team, 2015). We measured the predicted climatic niche overlap between pairs of lineages by calculating Schoener's D statistics from probability surfaces of the climatic niche models obtained from Maxent as described by Warren, Glor, and Turelli (2008). D ranges from 0 if niche models have no overlap to 1 if niche models fully overlap. To obtain a clustering representation of the difference in lineage climatic niche, we used the UPGMA method from "stats" 3.3.0 R package (R Development Core Team, 2015) in R.

We tested our models for niche identity based on D statistics, that is, whether the niches of each pair of lineages, as inferred by ENM, are more different than expected if they were drawn from the same underlying distribution (Warren *et al*., 2008). To test

niche identity for each pair of lineages X and Y, pseudoreplicate datasets were created by randomly partitioning the pooled set of X and Y occurrences, maintaining the same number of samples as those originally available for each lineage. A niche model was created from each pseudoreplicate, and its corresponding D statistic (De) was obtained. We used 100 replications to create a null distribution of similarity values. The observed values of D were compared to the percentiles of these null distributions of De in a one-tailed test to evaluate the null hypothesis that the niche models are not statistically significantly different. If D falls in the 95% of the null distribution, the null hypotheses that niche models are identical should not be rejected. Of course, tests of niche identity between allopatric species are not fully meaningful, as climatic niches of species with allopatric distributions are expected to differ, especially in areas like the Iberian Peninsula where different regions experience very different climatic conditions.

Finally, we tested for background similarities, that is, to determine whether ENMs are more similar than expected by chance, even if they occur in different regions (Warren *et al*., 2008). This type of analysis partly overcomes the problem that allopatric species often inhabit regions with distinct distributions of environmental variables, hence generating distinct realized climatic niche even without a difference in habitat selection patterns. This test compares the actual similarity of the models (assessed by Schoener's D) between pairs of lineages X and Y with a null distributions of 100 expected similarities (Ds) between a randomized dataset for Y based on random occurrence points drawn from within the range occupied by Y, and the symmetric randomized dataset for X. We defined the range available for each lineage using a 50km buffer zone around the occurrence points. Since most of the study area is covered with presence records, these buffer zones should be a good approximation of actual distributions without creating extensive false continuous occurrence areas between isolated populations. If D falls in the 95% of confidence limits of Ds, the null hypotheses of background similarity should not be rejected. Rejection of the null hypothesis indicates that the niche models of two *Podarcis* lineages are more similar or different than would be expected by chance given the existing differences between the environment in their distribution, hence that the observed niche differentiation between them is a function of habitat selection and/or suitability and not an artifact of the underlying environmental differences between the habitats available to each lineage.

To estimate the degree of geographic overlap of pairs of *Podarcis* lineages to understand whether niche divergence is accompanied, or not, by distinct geographic distributions, we first described the area of occurrence of each lineage and then calculated an index of pairwise geographic overlap. Because for some lineages we still lack accurate distribution limits, we described the current distribution range of each lineage by defining a 50 km buffer zone around the corresponding occurrence points, which should be a good approximation of actual distributions, as described before. We measured the distribution similarity of pairs of lineages using the binary Jaccard's similarity index (JSI) as implemented in the package "vegan" (Oksanen *et al*., 2017) for R. JSI ranges from 0, when two forms have no distribution grid squares in common, to 1 when all grid squares are shared. As for D, we used the UPGMA algorithm to obtain a clustering representation of global similarities in lineage distributions. To test for correlations between both D, or JSI of geographic distribution, and the genetic distance, we performed two Mantel tests. To calculate the genetic distances between each pair of lineages, we used the most complete genetic dataset available in terms of mtDNA fragment length: the dataset used in Kaliontzopoulou *et al*. (2011). We calculated Tamura and Nei (1993) genetic distance between each individual using ape 4.1 R package (Paradis, *et al*., 2004), and then, we used the average genetic distance between each pair of lineages to compute the Mantel test with ade4 1.7-6 R package (Dray & Dufour, 2007) to test the null hypothesis that niche overlap, or geographic distribution, and genetic distance are not related, with 9,999 permutations for $\alpha = 0.05$.

Results

Accuracy of the ecological niche models and variable importance

The Maxent model results presented in this study are the results obtained with sampling bias correction as both results are similar (see the subsection "Some methodological remarks" in "Discussion" for further details). We obtained consensus Maxent models with very high average AUC values (0.93–0.99, Table 2.1.). ENFA models also performed well, as indicated by a very high continuous Boyce index for most of models (0.79–0.99, Table 2.1.). Only the ENFA models for *P. hispanicus* Albacete/Murcia and Tunisia/Algeria group had lower (0.49 and 0.25, respectively) but still positive values, meaning that ENFA models are good.

For all lineages, at least one of the three variables related to precipitation (annual precipitation, precipitation in the driest month, and precipitation seasonality) had a considerable importance for the final Maxent models (21%–71%; Figure 3). For four lineages, only one variable had a very high importance (more than 50%): precipitation in the driest month for *P. hispanicus* Albacete/Murcia and *P. liolepis*; annual precipitation for Tunisia/Algeria group; and altitude for Jbel Siroua. Generally, temperature-related variables had lower contributions to the models (Figure 2.3.). The gains of most Maxent models considering each variable alone were higher for precipitation in the driest month, but it is evident the gain for Jbel Siroua when considering altitude and a minimum temperature of coldest month (Appendix I, Figure S1.1.). However, when excluding only one variable at a time, the gain was quite similar across variables for all models. Only when excluding annual precipitation for Tunisia/Algeria group was the gain lower (Appendix I., Figure S1.2.). Therefore, this variable provided exclusive information to the niche model of this lineage.

Table 0.1. Accuracy for average Maxent and ENFA models given by area under the curve (AUC) values and Boyce index (B), respectively, for each of the examined lineages of the *Podarcis hispanicus* complex and number of records for each one (N). Species acronyms as in Figure 1.

	Model												
	Pb	**Pc**	**Pgl**	**Pgg**	**Pv**	**Pl**	**PhG**	**PhAM**	**V**	**PvSS**	**PvMA**	**TAg**	**JS**
N	40	23	44	19	72	52	16	7	68	29	52	14	9
AUC	0.99	0.99	0.97	0.97	0.93	0.95	0.96	0.94	0.98	0.95	0.97	0.93	0.99
B	0.97	0.79	0.97	0.84	0.99	0.98	0.89	0.49	0.90	0.88	0.97	0.25	0.84

Variables response curves for each lineage are represented in detail in Appendix I., Figure S1.3. The probability of occurrence of almost all lineages is higher with medium levels of precipitation in the driest month and with moderate levels of precipitation seasonality, but the presence of *P. bocagei*, *P. carbonelli*, *P. g. lusitanicus*, and Tunisia/Algeria group is also more probable with high levels of annual precipitation. With some exceptions, the probabilities of occurrence of most lineages are lower with high maximum temperatures of the warmest month and low minimum temperatures of the coldest month. For Jbel Siroua, the probability of occurrence is very high with low values of minimum temperature of the coldest month and high altitudes.

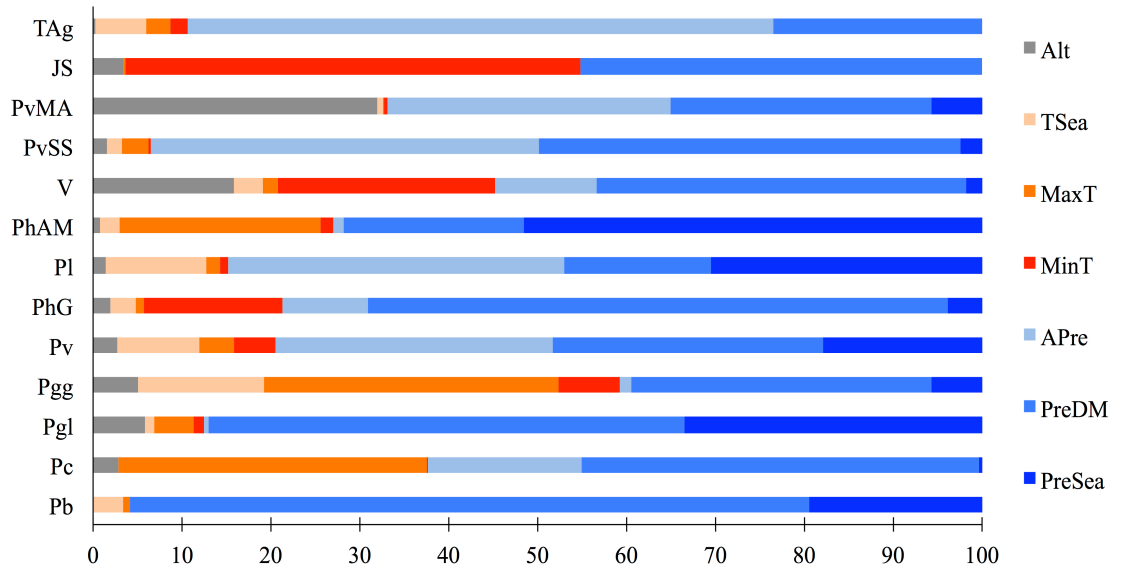

Figure 2.3. Contribution of each variable for the final Maxent models. Altitude (Alt), temperature seasonality (TSea), maximum temperature of the warmest month (MaxT), minimum temperature of the coldest month (MinT), annual precipitation (APre), precipitation of the driest month (PreDM), and precipitation seasonality (PreSea). Species acronyms as in Figure 1.

Examination of marginality versus tolerance values calculated by ENFA models (Figure 2.4.) showed a general trend toward low tolerances, even for those forms inhabiting wider areas, such as *P. liolepis*, *P. virescens*, and *P. vaucheri* Morocco/Algeria, and a continuum trend from low to high marginality. The lineage with the highest tolerance was *P. vaucheri* Morocco/Algeria, although its value can be considered relatively low (0.35). Jbel Siroua and *P. hispanicus* Albacete/Murcia were the lineages with the lowest tolerance values, very close to zero. Jbel Siroua was also the lineage with the highest marginality (1.68), while *P. hispanicus* Galera had the lowest marginality (0.14). Still, we can identify a group of lineages occupying average habitat conditions with marginality lower than 0.5: *P. hispanicus* Galera, *P. virescens*, Valencia, *P. g. guadarramae*, and *P. hispanicus* Albacete/Murcia; a group of forms occupying marginal habitats in the study area with marginality higher than 1.0: Jbel Siroua (the highest value), *P. bocagei*, *P. carbonelli*, and *P. g. lusitanicus*; and an intermediate group with marginality between 0.5 and 1.0 comprising Tunisia/Algeria group, *P. vaucheri* Southern Spain, P. vaucheri Morocco/Algeria, and P. liolepis.

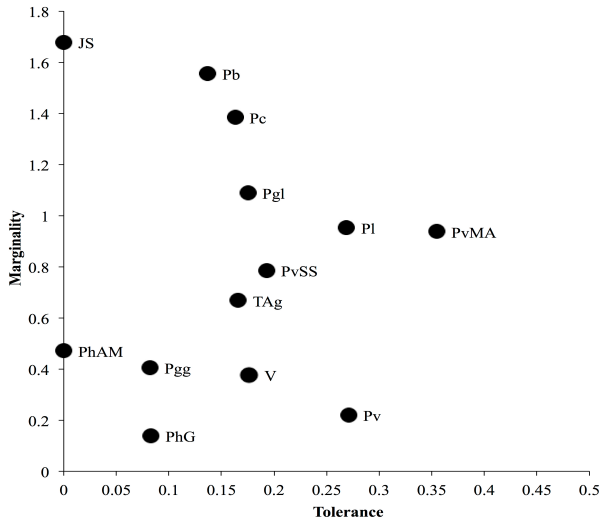

Figure 2.4. Marginality and Tolerance scores derived by ecological niche factor analysis for each lineage from the *Podarcis hispanicus* complex. Species acronyms as in Figure 2.1.

Predicted areas of suitability

According to Maxent results (Figure 2.5.), some models occupy Central and Southern Iberian Peninsula (*P. g. guadarramae*, *P. virescens*, *P. hispanicus* Galera, *P. vaucheri* Spain and *P. hispanicus* Albacete/ Murcia, Tunisia/Algeria group) as well as North of Africa. Lineages occurring in the north and west of the Iberian Peninsula (P. bocagei, *P. carbonelli*, P. g. lusitanicus, P. liolepis) and the two North African lineages mostly distributed in mountain regions (Jbel Siroua and *P. vaucheri* Morocco/Algeria) had more restricted predicted distributions (Figure 2.5.), mostly identical with the current known distributions. Detailed maps with each model with the presence records depicted are represented in Appendix I., Figure S1.4.

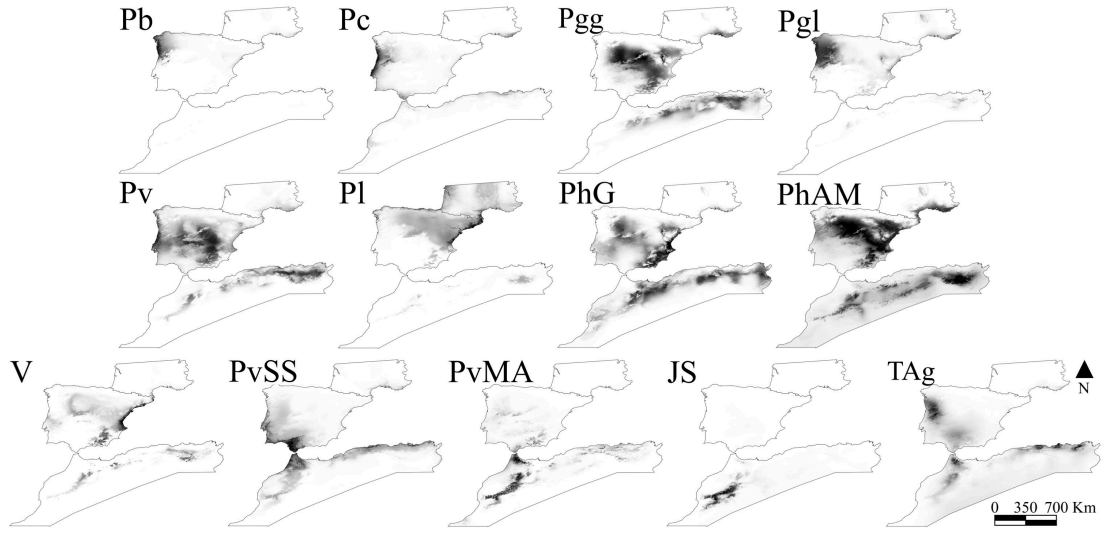

Figure 2.5. Maxent models for each form of *Podarcis hispanicus* complex. Dark and light colors represent areas of high and low suitability, respectively. Species acronyms as in Figure 2.1. For a detailed representation of the models with the presence records, see Figure S1.4.

Niche divergence in environmental and geographic spaces

The dendrogram constructed based on realized climatic niche divergence (D, Figure 2.6.a) indicated the existence of four clusters, but even within groups, we found high divergence levels. Jbel Siroua, *P. vaucheri* Morocco/Algeria, *P. vaucheri* South Spain, and Tunisia/ Algeria group were the first cluster to be segregated. Then, we can distinguish a cluster with *P. carbonelli*, *P. bocagei*, and *P. g. lusitanicus*. Finally, a group established by Valencia lineage, P. hispanicus Galera, P. g. guadarramae, and *P. virescens* was distinct from another cluster that includes *P. liolepis*, *P. hispanicus* Albacete/Murcia. Such results contrast with mtDNA phylogenetic relationships (Figure 2.2.) and estimates of relationships based on geographic distribution areas (JSI, Figure 2.6.b; see Appendix I., Table S1.2. for detailed results), which shows that most lineages have unique distributions, except for the two lineages from Northwest Iberia (*P. bocagei* and *P. g. lusitanicus*) and the three from southeast (Valencia lineage, *P. hispanicus* Galera, and *P. hispanicus* Albacete/Murcia).

Niche overlap D varied between 0.015 and 0.636, but most pairs exhibited a restricted niche overlap (Table 2.2.) as expected from their mostly allopatric distribution (Figure 6b). Of 78 pairs, only six had D > 0.5. When tested for niche identity, 66 of 78 pairs were significantly different (bold values in Table 2.), and thus, the null hypothesis of identical niche could be rejected in these cases. Niche background similarity tests revealed that 45 pairs were less similar than expected based on random sampling of regional differences in environmental variables (gray shaded values in Table 2.). For 33 comparisons, D did not significantly deviate from the expected distribution. No pairs of *Podarcis* lineages were more similar than what expected by chance. The results from the Mantel tests show no significant correlation between niche overlap and genetic distances (r = -0.206, p-value = 0.914) or geographic distribution and genetic distances (r = 0.055, p-value = 0.280).

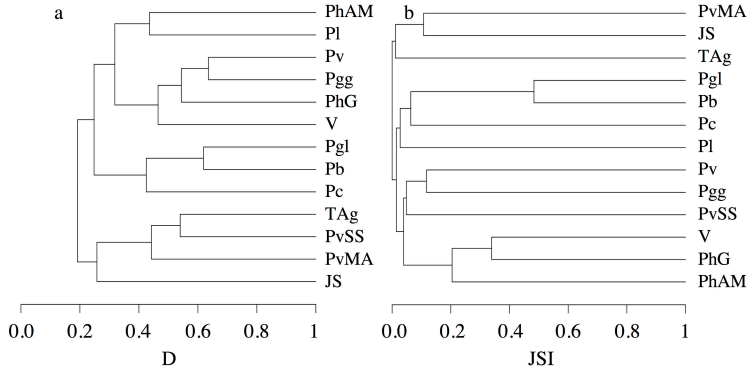

Figure 2.6. Dendrograms representing **a)** climatic niche similarity based on niche overlap (D) and **b)** geographic range similarity based on Jaccard's similarity index (JSI) among the *Podarcis hispanicus* complex. In both measures, 0 means no overlap and 1 complete overlap. Species acronyms as in Figure 2.1.

Table 0.2. Pairwise niche overlap (D) between each *Podarcis hispanicus* complex form and results for significant niche identity tests (bold) and niche background similarity tests (gray shaded).

	Pb	Pc	Pgl	Pgg	Pv	Pl	PhG	PhAM	V	PvSS	PvMA	JS
Pc	**0.344**											
Pgl	0.619	0.506										
Pgg	**0.141**	**0.258**	**0.332**									
Pv	**0.178**	**0.370**	**0.338**	0.636								
Pl	**0.250**	**0.163**	**0.306**	**0.324**	**0.301**							
PhG	**0.121**	**0.272**	**0.274**	0.576	0.512	**0.222**						
PhAM	**0.250**	**0.163**	**0.324**	**0.306**	**0.301**	0.436	**0.222**					
V	**0.167**	**0.228**	0.328	0.483	**0.493**	**0.436**	0.419	**0.435**				
PvSS	**0.111**	0.458	**0.192**	**0.231**	**0.402**	**0.109**	**0.304**	**0.109**	**0.191**			
PvMA	**0.081**	**0.301**	**0.184**	**0.308**	**0.351**	**0.115**	**0.316**	**0.037**	**0.299**	0.469		
JS	**0.015**	**0.083**	**0.044**	**0.140**	**0.111**	**0.037**	**0.187**	**0.067**	**0.102**	**0.157**	**0.399**	
TAg	**0.163**	**0.418**	**0.279**	**0.222**	**0.359**	**0.067**	**0.268**	**0.115**	**0.136**	0.540	0.416	**0.217**

Discussion

We used ecological niche modeling in the *Podarcis hispanicus* species complex to examine which climatic factors influence species' distributions and to assess how climatic niche divergence may contribute to lineage divergence and shape the geographic distribution of these lineages. We documented a high variation in niche similarity between pairs of lineages, ranging from almost no similarity (D = 0.015) to important niche similarity (D = 0.636), but without congruence between niche divergence, or geographic overlap, and mtDNA divergence. These results may be attributed to multiple events of allopatric divergence and shed more light on the ecological processes potentially involved in shaping lineage divergence and geographic space occupancy patterns in this group. Indeed, this lack of congruence may be the result of the divergence of closely related lineages in markedly distinct environments, while other more closely related lineages diverged in allopatric regions but with more similar environmental conditions.

Biogeographic and ecological affinities

The overall low niche tolerance observed across lineages suggests that *Podarcis* lineages have some degree of climatic specialization, and their marginality values follow a gradient of habitat from marginal to average conditions (Figure 2.4.). Low marginalities correspond mostly to forms occupying Mediterranean conditions (e.g., *P. hispanicus*

Galera, *P. g. guadarramae*, or *P. virescens*), while high marginalities correspond to Atlantic or mountainous conditions (e.g., *P. bocagei*, *P. g. lusitanicus*, or Jbel Siroua lineage). Such results point to distinct realized climatic niches that are, however, not completely divergent between several pairs of lineages.

Variables related to precipitation generally have the highest contribution to the models but their influence on distribution varies among lineages. For almost all lineages, we found that environmental suitability responds to precipitation in the driest month in a very similar way, with a peak of suitability around similar values of the lower end of the distribution of values for this variable, suggesting that all lineages select areas with similar and moderate values of rain in the driest month. This is not surprising as in general *Podarcis* lizards occupy areas in the study region with two well demarcated seasons, a humid season with higher levels of precipitation and a dry season with lower levels of precipitation more prominent across Mediterranean climatic regions. Still, species like *P. bocagei*, *P. g. lusitanicus* and *P. liolepis* exhibit larger standard deviations in the response curve for this variable (Appendix I., Figure S1.2.), as they can also occupy regions with higher precipitation levels. However, high values of annual precipitation seem to be important for *P. bocagei*, *P. carbonelli*, *P. g. lusitanicus*, *P. liolepis*, and Tunisia/Algeria group. Preference for habitats with relatively high humidity has been previously reported for several *Podarcis* species (e.g., *P. cretensis*, Herkt, 2007), including some of the lineages included in this study (e.g., *P. carbonelli*, Sá-sousa, 2001; Román *et al*., 2006; Sillero & Carretero, 2013; north African *Podarcis*, Kaliontzopoulou *et al*., 2008). High maximum temperatures during the warmest month hinder the occurrence of most lineages, for example, *P. liolepis*, *P. hispanicus* Galera, *P. bocagei*, *P. carbonelli*, *P. virescens*, Valencia, *P. vaucheri* Spain. Indeed, *Podarcis* lizards are known to avoid extremely arid, hot, and dry regions associated with desert environments in North Africa, which is most likely linked to the origin of the genus in more temperate Mediterranean environment (see Harris *et al*., 2002a; Harris, *et al*., 2002b; Lima *et al*., 2009). *Podarcis bocagei*, *P. carbonelli*, and *P. g. lusitanicus* seem more related to less marked temperature seasonality. This is congruent with a previous study where *P. bocagei* and *P. carbonelli* were associated with Atlantic bioclimatic regions, characterized by mild summers (Sillero *et al*., 2009).

The six lineages with the lowest marginality, that is, inhabiting environments closest to the average conditions available (Mediterranean climate), have high

proportions of predicted suitable niches both in the Iberian Peninsula and in North Africa (*P. hispanicus* Galera, *P. g. guadarramae*, *P. virescens*, *P. hispanicus* Albacete/Murcia, *P. vaucheri* Spain, and Tunisia/Algeria group). Because these lineages are not monophyletic and their most recent common ancestor is the ancestor of the entire group, it is likely that a niche related to Mediterranean climate is the ancestral state, while specialization to Atlantic or Atlantic-like marginal climates evolved later, independently.

Climatic niche divergence in a phylogenetic context

Our data did not reveal any relationships between climatic niche divergence and mtDNA divergence as shown by Mantel test and by the comparison of maximum likelihood estimates of mtDNA relationships and niche overlap among *Podarcis hispanicus* complex (Figures 2 and 6a). Despite methodological differences, several examples of closely related taxa that diverge greatly in niche have been revealed in previous studies (e.g., on dendrobatid frogs, Graham *et al*., 2004; Anolis sagrei group, Knouft *et al*., 2006). Several *Podarcis* sister mtDNA lineages have distinct distributions as shown by low JSI and no significant niche similarities (e.g., Jbel Siroua with Tunisia/ Algeria group, JSI = 0, D = 0.217; and *P. carbonelli* with *P. virescens*, JSI = 0.028, D = 0.370; divergence from 4 to 5.5 Mya, Kaliontzopoulou *et al*., 2011). This may be attributed to allopatric divergence in distinct environments. It is thus possible that significant realized climatic niche divergence occurred associated with several differentiation events under the influence of local adaptation (Warren *et al*., 2008). On the other hand, some highly divergent non-sister lineages with distinct distributions do not exhibit significant realized climatic niche differences (e.g., *P. hispanicus* Galera with *P. g. guadarramae*, JSI = 0; D = 0.576–10.4 Mya to the most recent common ancestor, Kaliontzopoulou et al., 2011). Convergence or long-term niche conservatism are both possibilities to explain such patterns (Knouft *et al*., 2006). Furthermore, some allopatric sister lineages without significant realized climatic niche differences (*P. vaucheri* Spain with *P. vaucheri* Morocco/Algeria, JSI = 0, D = 0.469) reflect cases with substantial environmental space overlap. Since similar conditions exist on both sides of the Strait of Gibraltar, these lineages have diverged in allopatry but under more or less similar environmental conditions.

By conducting analyses of background similarity, we found that 58% of pairs of lineages' realized climatic niche were more different than expected by chance because

the regions they occupy are environmentally different. This result suggests that the observed realized climatic niche differentiation between these pairs is partly due to different patterns of habitat selection within the heterogeneous environmental background. Ecological differentiation between these forms may reflect adaptation to different conditions driven by the differences in the heterogeneity within each environmental background (Warren *et al*., 2008; Warren *et al*. 2014). Parapatry between *P. bocagei* and *P. liolepis* in Northern Spain, for example, may thus be maintained by distinct environmental conditions. However, an important proportion of the models (42%) were no more similar or different than expected by chance.

Overall, our results are compatible with the hypothesis that genetic divergence across this group of lizards was likely built in allopatric conditions. In such a scenario, the almost continuous distributions (either in parapatry or partial sympatry) observed at present across this clade would have developed after secondary geographic contact between lineages. Although the analyses presented here cannot fully disclose whether realized climatic niche changes directly caused differentiation or speciation events, or whether alternatively are a consequence of lineage divergence, it is clear that realized climatic niche divergence is poorly related to genetic divergence as even closely related lineages have restricted realized climatic niche overlap.

Climatic niche divergence and geographic distribution

Identity tests reveal that most pairs of lineages (85%) are significantly different in the environmental space. Most of these pairs are formed at least by one lineage with low tolerance but high marginality (e.g., *P. hispanicus* Jbel Siroua, *P. bocagei*, *P. carbonelli*, *P. g. lusitanicus*), that is, lineages with high specialization, occurring in marginal environmental conditions in the study area, and with restricted geographic distributions. Because these lineages have no significant realized climatic niche similarities with most other lineages (see Table 2.2.), we can infer that their current distinct distribution ranges (Figure 6.b) may be mostly shaped by environmental conditions.

According to prediction maps of suitability (Figure 2.5.) and based also on the results of equivalence tests (Table 2.2.), considerable realized climatic niche overlap would be expected between some lineages (15%), which should exhibit extensive areas of overlap. Some of these lineages have high tolerance (e.g., *P. vaucheri* Morocco/ Algeria, *P. virescens*), reflecting large distribution ranges relatively to the study area, but

most are lineages with low tolerance and marginality (e.g., *P. hispanicus* Albacete/Murcia, Valencia, *P. hispanicus* Galera, *P. g. guadarramae*), indicating that they inhabit relatively small areas close to average conditions in this region. This last group of lineages tends to have the highest realized climatic niche similarities with several other lineages (see Table 2.2.). However, extensive areas of sympatry do not exist (Figure 6.b). On the other hand, *P. bocagei* and *P. g. lusitanicus* co-occur in extensive sympatric areas across their distributions, and *P. carbonelli* is everywhere sympatric with other *Podarcis* species.

Assuming that most *Podarcis* lineages have been well sampled, at least in the Iberian Peninsula and in Morocco, the lack of extensive overlap shows that factors other than bioclimatic and topographic variables are important in shaping range limits in this species complex. This is particularly evident for lineages with low tolerance and marginality, as they present restricted distributions but niche models tend to occupy relatively large geographic areas. Such results show that generally lineages' distributions may be primarily limited by physiological tolerances (Kellermann *et al*. , 2009), but a role for competition after secondary contact may be suggested, for example, between pairs like *P. virescens* and *P. g. guadarramae* or *P. hispanicus* Galera and *P. hispanicus* Albacete/Murcia. Dispersal constraints and/or historical factors may also be important limitations for their presence and the occurrence of extensive contact zones (Barbosa *et al*., 2012; Peterson, 2011). This may be illustrated by some cases where important geographic barriers are present or where geographic segregation is a consequence of the presence of other *Podarcis* lineages which diverged largely in allopatry (Diamond, 1973; Price, 2008). The first scenario includes *P. vaucheri* Southern Spain plus *P. vaucheri* Morocco/Algeria or *P. vaucheri* Southern Spain plus Tunisia/Algeria group disjointed by the Strait of Gibraltar. Actually, such allopatric lineages separated by the Strait of Gibraltar does not have significant realized climatic niche differences or significant distinct backgrounds as similar ecological conditions can be found in the Iberian Peninsula and North of Africa. In the second case, *P. virescens* occurrence between *P. hispanicus* Galera and *P. g. guadarramae* at central and southeastern Iberia may contribute to allopatry between these two latter lineages according to the distribution and overlap of suitable climatic areas for their occurrence. Jbel Siroua is in a similar situation with Tunisia/Algeria group and *P. vaucheri* Morocco/ Algeria, occupying most of the available North African habitats. In the last case, these forms are generally

allopatric, but the presence of *P. vaucheri* Morocco/Algeria may have contributed to a restricted geographic distribution as well as a more specialized realized climatic niche of the other two *Podarcis* lineages (Lima *et al*., 2009).

The almost continuous distribution of *Podarcis* lineages studied here results from a wide range of realized climatic niche divergence across the group. Furthermore, realized climatic niche overlap between several lineages suggests the potential for the existence of extensive areas of sympatry between several lineages that are, however, currently allopatric or parapatric. Hence, biotic interactions and historical factors seem to play an important role in shaping current patterns of distribution. Different realized climatic niches and competition both seem to interact to shape lineage distribution and limit overlap.

Some methodological remarques

Because it is still unsafe to identify all lineages of the *P. hispanicus* species complex on the basis of morphology (i.e., Kaliontzopoulou *et al*., 2012), the use of mtDNA for identification guarantees numerous genetically confirmed presence data. However, a certain level of mismatch with evolutionary relationships has to be expected, especially in the southeast of the Iberian Peninsula, where the Valencia lineage has been identified only as an introgressing mtDNA lineage in populations of *P. liolepis* or *P. hispanicus* Galera (Renoult *et al*., 2009).

Characterizing only a subset of the species fundamental niches leads to the limited representation of the full dimensions of these niches, that is, the realized niche (Peterson et al., 2011). Bioclimatic models can reconstruct environmental correlates of species geographic distributions (see Araújo & Peterson, 2012). Because all organisms require a stable input of energy for successful growth, survival, and reproduction (Porter & Gates, 1969; Kearney & Porter, 2004), the fundamental niche includes those environmental variables that influence organisms energy. Particularly for terrestrial ectotherms, like *Podarcis hispanicus* lineages, this “climate-space” is important. Since in this study, rather than predicting limits and detailed geographic distributions, we wish to understand environmental correlates that explain species distributions, and given the availability of spatiotemporal data for climatic conditions, the representation of the realized climatic niche obtained thus explains much about *Podarcis hispanicus* lineage

distributions as well as provides a solid basis from which to consider the effects of other environmental variables such as biotic interactions (Kearney & Porter, 2004).

An important issue was whether our dataset was geographically biased. To overcome this possible problem, we used a "restricted background" correction method proposed by Phillips (2008) and discussed by Fourcade *et al*. (2014). We are aware that this correction method may not be the most efficient in some cases (Fourcade *et al*., 2014), but after model calculations, the AUC values were the same for all lineages as the models without correction for sampling bias. So the models with this correction method did not perform worse or better as the uncorrected dataset. These results together with the fact that the total area where different lineages of *P. hispanicus* complex occur is well covered with samples, the sampling bias should not be a major concern and that general conclusions are identical if we use the uncorrected dataset.

Furthermore, despite the high resolution of the GIS data (1 km^2), the scale used in this study limits analyses to regions, rather than specific localities. We can only investigate environmental determinants of regional co-occurrence or allopatry, rather than strict sympatry or allopatric distribution at the local scale as differential habitat selection patterns may be responsible for the allopatric distribution at such scale. The next step should be the integration of fine-scale niche characteristics to elucidate the determinants of local distributions and habitat use, especially factors contributing to the formation and location of hybrid zones, which is of cornerstone importance for our understanding of the ecological and evolutionary processes taking place when different, closely related species, meet.

Acknowledgements

We thank P. Bacquet, O. Chaline, A. Cluchier, V. Fradet, M. Geniez, P. Geniez, C. P. Guillaume, G. Guillaume, J. Magraner, U. Mathis, V. Rufray, M. Siol, I. Rocha, D. Salvi, M. A. Carretero, C. Rato, A. Perera, M. Ribeiro, D. Donaire, F. Jorge, G. Perez-Lanuza, E. Ayllón, P. Hernández- Sastre, J. C. Barberá, M. Barata, F. Martínez-Freiría, D. J. Harris, J. Maia, B. Tomé, I. Damas and J. C. Brito for helping with sample collection, P. Bacquet, M. Beddek, P. Sourrouille and M. Ribeiro for sequencing some of the samples and P. Geniez for managing the tissue database in Montpellier and helping us in many other ways. GCD was supported by a PhD grant (SFRH/BD/89750/2012); CL by a

research grant (PTDC/BIA-BEC/102179/2008); NS, CP, and AK by IF contracts (IF/01526/2013, IF/01597/2014/CP1256/ CT0009, and IF/00641/2014/CP1256/CT0008, respectively), all under the Programa Operacional Potencial Humano—Quadro de Referência Estratégico Nacional funds from the European Social Fund and Portuguese Ministério da Educação e Ciência. Support was also provided by FCT project PTDC/BIA-BEC/102179/2008, under FEDER COMPETE funds (FCOMP-01-0124-FEDER-007062) and by the project "Biodiversity, Ecology and Global Change" co-financed by North Portugal Regional Operational Program 2007/2013 (ON.2—O Novo Norte), under the National Strategic Reference Framework (NSRF), through the European Regional Development Fund (ERDF).

References

Araújo MB, New M (2007) Ensemble forecasting of species distributions. *Trends in Ecology & Evolution*, **22**, 42–47.

Araújo MB, Peterson, A. T. (2012) Uses and misuses of bioclimatic envelope modeling. *Ecology*, **93**, 1527–1539.

Arnold EN, Arribas O, Carranza S (2007) Systematics of the palaearctic and oriental lizard tribe Lacertini (Squamata: *Lacertidae*: *Lacertinae*), with descriptions of eight new genera. *Zootaxa*, **1430**, 1–86.

Arntzen JW, Sá-Sousa P (2007) Morphological and genetical differentiation of lizards (*Podarcis bocagei* and *P. hispanica*) in the Ria de Arosa archipelago (Galicia, Spain) resulting from vicariance and occasional dispersal. In W. Renema (Ed.), Biogeography, time and place: distributions, barriers, and islands (pp. 365–401) Dordrecht: Springer.

Barbosa MA, Sillero N, Martínez-Freiría F, Real R (2012) Ecological niche models in Mediterranean herpetology: past, present and future. In W. Zhang, (Ed.), Ecological modeling (pp. 173–204) Hauppauge (NY): Nova Publishers.

Boyce MS, Vernier PR, Nielsen SE, Schmiegelow FK (2002) Evaluating resource selection functions. *Ecological Modelling*, **157**, 281–300. https://doi.org/10.1016/S0304-3800(02)00200-4

Busack SD, Lawson R, Arjo WM (2005) Mitochondrial DNA, allozymes, morphology and historical biogeography in the *Podarcis vaucheri* (*Lacertidae*) species complex. *Amphibia-Reptilia*, **26**, 239–256. https://doi.org/10.1163/1568538054253438

Carranza S, Arnold EN, Amat F (2004) DNA phylogeny of Lacerta (*Iberolacerta*) and other lacertine lizards (Reptilia: *Lacertidae*): did competition cause long-term mountain restriction? *Systematics and Biodiversity*, **2**, 57–77.

Castilla AM, Fernández-Pedrosa V, Harris DJ, González A, Latore A, Moya A (1998) Mitochondrial DNA divergence suggests that *Podarcis hispanica atrata* (Squamata: *Lacertidae*) from Columbretes islands merits specific distinction. Copeia, **1998**, 1037–1040.

Coyne JA, Orr HA (2004) Speciation. Sunderland, MA. Diamond, J. M. (1973) Ensure survival of as many native Distributional Ecology of New Guinea Birds. *Advanced Science*, **179**, 759–769.

Dias G, Luis C, Pinho C, Kaliontzopoulou A (2016) A case of *Podarcis carbonelli* intake by *Podarcis virescens*. *Herpetology Notes*, **9**, 105–108.

Dray S, Dufour AB (2007) The ade4 package: Implementing the duality diagram for ecologists. *Journal of Statistical Software*, **22**, 1–20.

Elton C (1927) Animal ecology. London: Sedgwick and Jackson.

Fourcade Y, Engler JO, Rodder D, Secondi J (2014) Mapping species distributions with Maxent using a geographically biased sample of presence data: a performance assessment of methods for correcting sampling bias. *PLoS ONE*, **9**, e97122.

Geniez P, Cluchier A, Sá-Sousa P, Guillaume CP, Crochet PA (2007) Systematics of the *Podarcis hispanicus*-complex (Sauria, *Lacertidae*) I: Redefinition, morphology and distribution of the nominotypical taxon. *The Herpetological Journal*, **17**, 69–80.

Geniez P, Sá-Sousa P, Guillaume CP, Cluchier A, Crochet PA (2014) Systematics of the *Podarcis hispanicus* complex (Sauria, *Lacertidae*) III: Valid nomina of the western and central Iberian forms. *Zootaxa*, **3794**, 1–51.

Graham CH, Ron SR, Santos JC, Schneider CJ, Moritz C (2004) Integrating phylogenetics and environmental niche models to explore speciation mechanisms in dendrobatid frogs. *Evolution*, **58**, 1781–1793.

Grinnell J (1917) The niche-relationships of the California Thrasher. Auk, 34, 427–433.

Guisan A, Zimmermann NE (2000) Predictive habitat distribution models in ecology. *Ecological Modelling*, **135**, 147–186.

Harris DJ, Arnold EN (1999) Relationships of wall lizards, *Podarcis* (Reptilia: *Lacertidae*) based on mitochondrial DNA sequences. *Copeia*, **3,** 749–754.

Harris DJ, Batista V, Carretero MA, Pinho C, Sá-Sousa P (2002a) Mitochondrial DNA sequence data confirms the presence of *Podarcis carbonelli*, Pérez-Mellado, 1981 in southern Spain. *Herpetozoa*, **15**, 188–190.

Harris DJ, Carranza S, Arnold EN, Pinho C, Ferrand N (2002b) Complex biogeographical distribution of genetic variation within *Podarcis* wall lizards across the Strait of Gibraltar. *Journal of Biogeography*, **29**, 1257–1262.

Harris DJ, Sá-Sousa P (2001) Species distinction and relationships of the western Iberian *Podarcis lizards* (Reptilia, *Lacertidae*) based on morphology and mitochondrial D sequences. *Herpetology Journal*, **11**, 129–136.

Harris DJ, Sá-Sousa P (2002) Molecular phylogenetics of Iberian wall lizards (*Podarcis*): is *Podarcis hispanica* a species complex? Molec*ular Phylogenetics and Evolution*, **23**, 75–81.

Herkt M (2007) Modelling habitat suitability to predict the potential distribution of Erhard's wall lizard *Podarcis erhardii* on Crete (Master Thesis) Retrieved from http://www.podarcis.de/AS/Bibliografie/BIB_1440.pdf

Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A (2005) Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, **25**, 1965–1978.

Hirzel AH, Hausser J, Chessel D, Perrin N (2002) Ecological-niche factor analysis: how to compute habitat-suitability maps without absence data? *Ecology*, **83**, 2027–2036.

Hirzel AH, Hausser J, Perrin N (2006) Biomapper 4.0. Lab. of Conservation Biology, Department of Ecology and Evolution, University of Lausanne, http://www. unil. ch/biomapper.

Holt RD (2003) On the evolutionary ecology of species' ranges. *Evolutionary Ecology Research*, **5**, 159–178.

Hutchinson GE (1957) The multivariate niche. Cold Spring Harbor Symposia on Quantitative Biology, **2**, 415–421. https://doi.org/10.1101/ SQB.1957.022.01.039

Jackson ST, Overpeck J T (2000) Responses of plant populations and communities to environmental changes of the late Quaterry. *Paleobiology*, **26**, 194–220.

Kaliontzopoulou A, Brito JC, Carretero MA, Larbes S, Harris DJ (2008) Modelling the partially unknown distribution of wall lizards (*Podarcis*) in North Africa: ecological affinities, potential areas of occurrence, and methodological constraints. *Canadian Journal of Zoology*, **86**, 992–1001.

Kaliontzopoulou A, Carretero MA, Llorente GA (2012) Morphology of the *Podarcis* wall lizards (Squamata: *Lacertidae*) from the Iberian Peninsula and North Africa: patterns of variation in a putative cryptic species complex. *Zoological Journal of the Linnean Society*, **164**, 173–193.

Kaliontzopoulou A, Pinho C, Harris DJ, Carretero MA (2011) When cryptic diversity blurs the picture: a cautionary tale from Iberian and North African *Podarcis* wall lizards. *Biological Journal of the Linnean Society*, **103**, 779–800.

Kearney M, Porter WP (2004) Mapping the fundamental niche: physiology, climate, and the distribution of a nocturnal lizard. *Ecology*, **85**, 3119–3131.

Kellermann V, van Heerwaarden B, Sgrò CM, Hoffmann AA (2009) Fundamental evolutionary limits in ecological traits drive *Drosophila* species distributions. *Science*, **325**, 1244–1246.

Kidd DM, Ritchie MG (2006) Phylogeographic information systems: putting the geography into phylogeography. *Journal of Biogeography*, **33**, 1851–1865.

Knouft JH, Losos JB, Glor RE, Kolbe JJ (2006) Phylogenetic analysis of the evolution of the niche in lizard of the *Anolis sagrei* group. *Ecology*, **87**, S29–S38.

Lima A, Pinho C, Larbes S, Carretero MA, Brito JC, Harris DJ (2009) Relationships of *Podarcis* wall lizards from Algeria based on mtDNA data. *Amphibia-Reptilia*, **30**, 483–492.

Liu C, Berry PM, Dawson TP, Pearson RG (2005) Selecting thresholds of occurrence in the prediction of species distributions. *Ecography*, **28**, 385–393.

Mayr E (1963) Animal species and evolution. Cambridge, MA: Harvard University Press. https://doi.org/10.4159/harvard.9780674865327

Oksanen J, Blanchet FG, Friendly M, Kindt R, Legendre P, McGlinn D, Wagner H (2017) vegan: Community Ecology Package. R package version 2.4-4. https://CRAN.R-project.org/package=vegan.

Oliverio M, Bologna MA, Mariottini P (2000) Molecular biogeography of the Mediterranean lizards *Podarcis* Wagler, 1830 and *Teira* Gray, 1838 (Reptilia, Lacertidae) *Journal of Biogeography*, **27**, 1403–1420.

Paradis E, Claude J, Strimmer K (2004) APE: Analyses of phylogenetics and evolution in R language. *Bioinformatics*, **20**, 289–290.

Pearman PB, Lavergne S, Roquet C, Wüest R, Zimmermann NE, Thuiller W (2014) Phylogenetic patterns of climatic, habitat and trophic niches in a European avian assemblage. Global *Ecology and Biogeography*, **23**, 414–424.

Pearson RG (2007) Species' distribution modeling for conservation educators and practitioners: Synthesis. New York: American Museum of Natural History.

Pearson RG, Raxworthy CJ, Nakamura M, Peterson AT (2007) Predicting species distributions from small numbers of occurrence records: a test case using cryptic geckos in Madagascar. *Journal of Biogeography*, **34**, 102–117.

Pérez-Mellado V (1981) Nuevos datos sobre la sistemática y distribución de *Podarcis bocagei* (Seoane, 1884) (Sauria, Lacertidae) en la Península Ibérica. *Amphibia-Reptilia*, **2**, 259–265.

Peterson AT (2011) Ecological niche conservatism: a time-structured review of evidence. *Journal of Biogeography*, **38**, 817–827.

Peterson AT, Soberón J, Pearson RG, Anderson RP, Nakamura M, Martinez-Meyer E, Araújo MB (2011) Ecological niches and geographical distributions. Princeton University Press. Princeton, New Jersey, USA.

Peterson AT, Soberón J, Sanchez-Cordero V (1999) Conservatism of ecological niches in evolutionary time. *Science*, **285**, 1265–1267.

Phillips SJ (2008) Transferability, sample selection bias and background data in presence-only modelling: a response to Peterson *et al.* (2007) *Ecography*, **31**, 272–278.

Phillips SJ, Anderson RP, Schapired RE (2006) Maximum entropy modeling of species geographic distributions. *Ecological Modelling*, **190**, 231–259.

Phillips SJ, Dudík M (2008) Modeling of species distribution with Maxent: new extensions and a comprehensive evalutation. *Ecography*, **31**, 161–175.

Phillips SJ, Dudík M, Schapire RE (2004) A maximum entropy approach to species distribution modeling. Twenty-first international conference on machine learning - ICML '04, 83.

Pinho C, Ferrand N, Harris DJ (2006) Reexamination of the Iberian and North African *Podarcis* (Squamata: *Lacertidae*) phylogeny based on increased mitochondrial DNA sequencing. *Molecular Phylogenetics and Evolution*, **38**, 266–273.

Pinho C, Harrid DJ, Ferrand N (2007) Contrasting patterns of population subdivision and historical demography in three western Mediterranean lizard species inferred from mitochondrial DNA variation. *Molecular Ecology*, **16**, 1191–1205.

Pinho C, Harris DJ, Ferrand N (2007) Comparing patterns of nuclear and mitochondrial divergence in a cryptic species complex: the case of Iberian and North African wall lizards (Podarcis, Lacertidae) *Biological Journal of the Linnean Society*, **91**, 121–133.

Pinho C, Harris DJ, Ferrand N (2008) Non-equilibrium estimates of gene flow inferred from nuclear genealogies suggest that Iberian and North African wall lizards (*Podarcis* spp) are an assemblage of incipient species. *BMC Evolutionary Biology*, **8**, 63.

Pinho C, Kaliontzopoulou A, Carretero MA, Harris DJ, Ferrand N (2009) Genetic admixture between the Iberian endemic lizards Podarcis bocagei and *P. carbonelli*: evidence for limited natural hybridization and a bimodal hybrid zone. *Journal of Zoological Systematics and Evolutionary Research*, **47**, 368–377.

Porter WP, Gates DM (1969) Thermodynamic equilibria of animals with environment. *Ecological Monographs*, **39**, 227 244. https://doi.org/10.2307/1948545

Price T (2008) Speciation in birds. Greenwood Village, CO: Roberts. QGIS Development Team (2013) QGIS Geographic Information System. Open Source Geospatial Foundation Project. Retrieved from http://qgis.osgeo.org/en/site/

R Core Team (2013) R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org/

Renoult JP (2006) Histoire évolutive du Lézard ibérique *Podarcis hispanicus* (Squamata: *Lacertidae*): discordances entre morphologie, ADN nucléaire et ADN mitochondrial. Master 2 Recherche. Université Montpellier II Sciences et Techniques du Languedoc.

Renoult JP, Geniez P, Bacquet P, Benoit L, Crochet PA (2009) Morphology and nuclear markers reveal extensive mitochondrial introgressions in the Iberian Wall Lizard species complex. *Molecular Ecology*, 18, 4298–4315.

Rieseberg LH, Raymond O, Rosenthal DM, Lai ZK, Livingston K, Nakazato T, *et al.* (2003) Major ecological transitions in wild sunflowers facilitated by hybridization. *Science*, 301, 1211–1216.

Rissler LJ, Apodaca JJ (2007) Adding more ecology into species delimitation: ecological niche models and phylogeography help define cryptic species in the black salamander (*Aneides flavipunctatus*) *Systematic Biology*, 56, 924–942.

Román J, Ruiz G, Delibes M, Revilla E (2006) Factores ambientales condicionantes de la presencia de la lagartija de Carbonell Podarcis carbonelli (Pérez-Mellado, 1981) en la comarca de Doñana. *Animal Biodiversity and Conservation*, 29, 73–82.

Sanz-Azkue I, García-Etxebarria K, Gosá A, Rubio X, Jugo BM (2006) Phylogenetic description of the Iberian wall lizard (*Podarcis hispanica*) in Donostia-San Sebastián and neighbouring populations IX Congresso Luso-Espanhol, XXIII Congerso Español de Herpetologia, San Sebastián.

Sá-Sousa P (2001) Comparative chorology between *Podarcis bocagei* and *P. carbonelli* (Sauria: Lacertidae) in Portugal. Revista Española de Herpetologia, 15, 85–97.

Sá-Sousa, P, Harris, D. J. (2002) *Podarcis carbonelli* Perez-Mellado, 1981 is a distinct species. *Amphibia-Reptilia*, 23, 459–468.

Schluter D (2001) Ecology and the origin of species. *Trends in Ecology & Evolution*, 16, 372–380.

Schluter D (2009) Evidence for ecological speciation and its alternative. *Science*, 323, 737–741.

Sillero N (2011) What does ecological modelling model? A proposed classification of ecological niche models based on their underlying methods. *Ecological Modelling*, 222, 1343–1346.

Sillero N, Campos J, Bordi A, Corti C, Creemers R, Crochet PA, *et al.* (2014) Updated distribution and biogeography of amphibians and reptiles of Europe. *Amphibia-Reptilia*, 35, 1–31.

Sillero N, Carretero MA (2013) Modelling the past and future distribution of contracting species. The Iberian lizard *Podarcis carbonelli* (Squamata: *Lacertidae*) as a case study. *Zoologische Anzeiger*, **252**, 289–298.

Sillero N, Brito JC, Skidmore AK, Toxopeus AG (2009) Biogeographical patterns derived from remote sensing variables: the amphibians and reptiles of the Iberian Peninsula. *Amphibia-Reptilia*, **30**, 185–206. https://doi.org/10.1163/156853809788201207

Soberón J (2007) Grinnellian and Eltonian niches and geographic distributions of species. *Ecology Letters*, **10**, 1115–1123.

Soberón J, Nakamura M (2009) Niches and distributional areas: concepts, methods, and assumptions. *Proceedings of the National Academy of Sciences*, 106, 19644–19650.

Soberón J, Peterson AT (2005) Interpretation of models of fundamental ecological niches and species distributional areas. *Biodiversity Informatics*, **2**, 1–10.

Swenson NG (2006) GIS-based niche models reveal unifying climatic mechanisms that maintain the location of avian hybrid zone locations in a North American suture zone. *Journal of Evolutionary Biology*, **19**, 717–725.

Swenson NG, Fair JM, Heikoop J (2008) Water stress and hybridization between *Quercus gambelii* and *Quercus grisea*. *Western North American Naturalist*, **68**, 498–507.

Tamura K, Nei M (1993) Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Molecular Biology and Evolution*, **10**, 512–526.

USGS (2006) Shuttle Radar Topography Mission, 30 Arc Second scene SRTM_GTOPO_ u30_mosaic, Global Land Cover Facility, University of Maryland, College Park, Maryland, February 2000. Retrieved from http://glcf.umd.edu/data/srtm/

Warren DL (2012) In defense of "niche modeling". *Trends in Ecology & Evolution*, **27**, 497–500.

Warren DL (2013) "Niche modeling": That uncomfortable sensation means it's working. A reply to McInerny and Etienne. *Trends in Ecology & Evolution*, **28**, 193–194.

Warren DL, Cardillo M, Rosauer DF, Bolnick DI (2014) Mistaking geography for biology: inferring processes from species distributions. *Trends in Ecology & Evolution*, **29**, 572–580. https://doi.org/10. 1016/j.tree.2014.08.003

Warren DL (2016) Analysis of niche evolution using niche and distribution models. version 0.1. https://github.com/danlwarren/ENMTools.

Warren DL, Glor RE, Turelli M (2008) Environmental niche equivalency versus conservatism: quantitative approaches to niche evolution. *Evolution*, **62**, 2868–2883.

Wiens JJ (2004) Speciation and ecology revisited: Phylogenetic niche conservatism and the origin of species. *Evolution*, **58**, 193–197.

Wiens JJ, Graham CH (2005) Niche conservatism: integrating evolution, ecology, and conservation biology. *Annual Review of Ecology, Evolution and Systematics*, **36**, 519–553.

Wright S (1921) Correlation and causation. *Journal of Agricultural Research*, **20**, 557–585.

Chapter 3. Genetic basis of reproductive isolation between some species from *Podarcis hispanicus* complex

Article II. Genome wide patterns of interspecific admixture in a natural hybrid zone in late stages of speciation.

Caeiro-Dias G, Brelsford A, Ribeiro M, Crochet PA, Pinho C (in prep.)

Article III. Evolution of sympatry without complete reproductive isolation: does hybridization matter for *Podarcis carbonelli* conservation?

Caeiro-Dias G, Brelsford A, Crochet PA, Pinho C (in prep.)

Article II. Genome-wide patterns of interspecific admixture in a hybrid zone between two *Podarcis* wall lizards in late stages of speciation

Guilherme Caeiro-Dias[1,2,3], Alan Brelsford[4,5], Mariana Ribeiro[1], Pierre-Andre Crochet[3], Catarina Pinho[1]

[1]Centro de Investigação em Biodiversidade e Recursos Genéticos, CIBIO/InBio, Vairão, Portugal
[2]Departamento de Biologia, Faculdade de Ciências, Universidade do Porto, Porto, Portugal
[3]CEFE, CNRS, University of Montpellier, University Paul Valéry Montpellier 3, EPHE, IRD, Montpellier, France
[4]Department of Ecology and Evolution, University of Lausanne, Lausanne, Switzerland
[5]Biology Department, University of California Riverside, California, USA

Abstract

In organisms with sexual reproduction, speciation occurs when increasing divergence results in prezygotic or postzygotic reproductive isolation between sets of the population. One of the main mechanisms of postzygotic isolation is the build up of genetic incompatibilities resulting in reduced viability or fertility of hybrids. Genetic incompatibilities may not accumulate homogenously across the genome and thus when nascent species hybridize, introgression can affect individual loci differently depending on whether they are involved in incompatibilities and on their genomic context. Studies aiming to decipher whether reproductive isolation arises via genomic islands of differentiation or from loci scattered throughout the genome are typically hampered by two problems: first, they often rely on indirect measures of introgression like those obtained via genome scans; second they focus on the early stages of speciation, providing only limited evidence with respect to the genetic architecture that promotes the maintenance of reproductive isolation in the long term. In this study, we overcome both of these problems by focusing on direct measures of introgression in a pair of species in their late stages of speciation. We use RADseq genotyping in the only known natural hybrid zone between *Podarcis bocagei* and *P. carbonelli* to examine the extent of hybridization, level of admixture and variation in selection against introgression among

distinct regions of the genome. These two species were found to hybridize in their narrow zone of contact but most individuals were assigned to one of pure parentals, evidencing the bimodality of the hybrid zone and the existence of strong mechanisms of reproductive isolation. A geographic cline approach revealed the presence of strong barriers to gene flow that prevent extensive introgression outside of the contact zone, except for a few signals suggesting that this might be a moving hybrid zone. Genomic cline analysis revealed heterogeneous patterns of introgression among loci within the contact zone, often with patterns different from those expected given the geographic cline inferences. For example, only 25% of the genome was inferred to have restrictions to gene flow, which contrasts with the strong bimodality and sharp concordant geographic clines. By comparing the regions where SNPs were discovered with the closely related *Podarcis muralis* genome we found that that strong barriers to gene flow are widespread in the genome, but a particularly important role in reproductive isolation between these two species is suggested for the Z chromosome. These patterns of introgression thus seem to be maintained by a combination of intrinsic and extrinsic barriers to interspecific gene flow building strong reproductive isolation between both species.

Key words: reproductive isolation; geographic clines; genomic clines; *Lacertidae*; RAD sequencing.

Introduction

The evolution of reproductive isolation is a gradual process (Templeton 1992) and hybridization often persists from the initial stages of speciation (Mallet 2005; Rieseberg 2009) to more advanced stages (Nosil *et al.* 2009). Hybridization and incomplete reproductive isolation may thus result in introgressive gene flow for most of the duration of the speciation process (Mullen *et al.* 2008; Melo-Ferreira *et al.* 2009; Tarroso *et al.* 2014). However, as divergence time increases, the development of genetic incompatibilities impedes gene flow, gradually causing complete reproductive isolation via hybrid unviability or sterility (Dobzhansky 1937; Muller 1942; Gourbiere & Mallet 2010; Matute *et al.* 2010). Deciphering the nature, number and distribution of these

genetic incompatibilities across the genome remains one of the main challenges of speciation.

One hypothesis to explain the build-up of reproductive isolation during divergence is that the accumulation of such incompatibilities occurs in 'genomic islands' of reduced gene flow and elevated differentiation (Turner *et al.* 2005; Feder & Nosil 2010; Ellegren *et al.* 2012). Such genomic islands become more extensive when divergent selection reduces effective gene flow in the surrounding genomic regions, i.e. divergence hitchhiking (Feder & Nosil 2010; Via 2012), and may be particularly important for speciation with gene flow if blocks of associated loci that contribute to isolation are co-inherited (Feder *et al.* 2012a). An alternative hypothesis to explain the development of reproductive isolation is that gene incompatibilities arise in many unlinked loci scattered throughout the genome (Feder & Nosil 2010; Feder *et al.* 2012b). These two models are expected to produce distinguishable patterns of variation in introgression rate before the completion of the speciation process but ultimately they result in identical situations where genomes are too divergent to produce fertile hybrid offspring. Empirical studies using newly available genomic tools (Gompert *et al.* 2010; Ellegren *et al.* 2012; Parchman *et al.* 2013; Larson *et al.* 2014) support the idea that the genome is not a homogenous ensemble (Wu 2001; Nosil & Feder 2012) and that reproductive isolation affects individual loci differently depending on their linkage to incompatibility genes. Whether the genetic architecture of reproductive isolation in later stages is concordant to one model or the other, or even if it will be possible to distinguish them, remains to be fully evaluated. Several empirical studies have identified few restricted genomic regions with major consequences in adaptation and isolation (e.g. Lawniczak *et al.* 2010; Ellegren *et al.* 2012; Strasburg *et al.* 2012) as well as many small regions, whose number increases with time of divergence and strength of selection (Yatabe *et al.* 2007), and which are distributed throughout the genome (Nakazato *et al.* 2007; Michel *et al.* 2010).

Part of the difficulties in evaluating the relative importance of these models is related to the exclusively comparative framework in which most speciation genomics studies are accomplished. Typically such studies use indirect approaches like genomic scans to make inferences about introgression. Interpreting the peaks and troughs of differentiation in indirect approaches is not straightforward, as similar genomic patterns can be generated by several processes and/or influenced by multiple factors that vary

across the genome and during different stages of speciation (Cruickshank & Hahn 2014; reviewed in Ravinet *et al.* 2017). On the opposite, the study of hybrid zones and the collection of direct measures of introgression can help to overcome these issues (Harrison & Larson 2016). Genomic approaches applied to the analyses of hybrid zones allow quantifying admixture and introgression to identify putative genomic regions that contribute to reproductive isolation (Szymura & Barton 1986; Gompert *et al.* 2012b): loci that are involved in reproductive isolation (speciation loci) or that are physically linked to speciation loci should exhibit reduced introgression when compared to the rest of the genome. If speciation originates on the evolution of genomic islands of differentiation, we would expect to see two populations of loci with contrasting introgression patterns for most of the speciation process, with loci that introgress less grouped in few restricted portions of the genome. However, if speciation develops from many unlinked loci scattered throughout the genome or if it is close to completion, we would not expect dramatic variation in introgression patterns across the genome and no physical clumping of loci with reduced introgression.

A second problematic issue about the majority of studies on the genomics of divergence is their focus on the early stages of speciation (e.g. Teeter *et al.* 2010; Janoušek *et al.* 2012; Luttikhuizen *et al.* 2012; Andrew & Rieseberg 2013; Parchman *et al.* 2013). These early stages are clearly the best to document which are the initial genomic architectures promoting the onset of differentiation. However, the factors influencing genomic patterns may vary during different stages of speciation (Ravinet *et al.* 2017). Forms in their very early stages of speciation represent states that may not maintain reproductive isolation in the long term; as such, studying genomic mechanisms by which reproductive isolation is maintained – and not only initiated – is equally important, thus suggesting that the study of genomic divergence and gene flow in the later stages of speciation is also invaluable to understand which factors influence the genomic architecture of reproductive isolation across the whole process of speciation.

In this study, we take a direct approach to the study of introgression by quantifying locus-specific patterns in a hybrid zone. We focus on a pair of species in their late stages of speciation which are excellent models for this study: *Podarcis bocagei* and *P. carbonelli*. These two species are part of the *Podarcis hispanicus* complex (Iberian wall lizards) and, despite their deep divergence (Kaliontzopoulou *et al.* 2011), interspecific gene flow has been reported (Pinho *et al.* 2009). Previous studies in

this contact zone (Pinho *et al.* 2009; Ribeiro 2014) showed that the two species hybridize, resulting in a narrow and bimodal hybrid zone; however the low number of loci employed precluded fine examination of the genomic patterns of introgression.

Here we use RADseq genotyping in the natural hybrid zone between these two species to examine the stage where the two forms lie in the speciation continuum and evaluate genomic patters of introgression. We first assess the strength of reproductive barriers by performing a simulation study. Then we evaluate the proportion of the genome that might be involved in reproductive isolation and the nature of possible barriers to interspecific gene flow. We used two complementary approaches to detect variation in introgression levels among loci. Finally, we mapped our SNPs on the recently available genome of *Podarcis muralis* (Andrade *et al.* submitted) to identify whether the loci that show reduced introgression are spread over the genome or grouped in small regions corresponding to islands of differentiation as well as to assess the role of the sexual chromosome in the reproductive isolation between *P. bocagei* and *P. carbonelli*.

Materials and methods

Transect sampling

From April to September 2013, we collected tissue samples across a 44 km transect along the north-western coast of Portugal, crossing the narrow contact zone between *P. bocagei* and *P. carbonelli* (Figure 3.1.1. a and b). This contact zone spans a few hundred meters in a narrow coastal dune stripe north of Espinho (41.020 N / 8.643 W, Aveiro District, Portugal). In Espinho (ESP), the population previously identified as the hybrid zone (Pinho *et al.* 2009), we sampled 161 individuals. We also sampled lizards in eight locations between the city of Porto in the north and the village of Torreira in the south (Figure 3.1.1.b). South of ESP we collected *carbonelli* individuals in Silvalde (SIL, n=32), Esmoriz (ESM, n=26) and Torreira (TOR, n=23); north of ESP we collected *bocagei* in Aguda (AGU, n=23), Francelos (FRA, n=18), Madalena (MAD, n=26) and Porto (POR, n=20). Note that no area of sympatry is known outside ESP and that only *P. bocagei* occurs north of ESP and only *P. carbonelli* south of ESP (Pinho *et al.* 2009). Lizards were captured by noose around the neck, a harmless capture method, and kept in individual cloth bags until processed. All the specimens were geo-referenced and

photographed. A small tail tip was collected and immediately stored in 96% ethanol. Animals were released in the place of capture the same day they were caught.

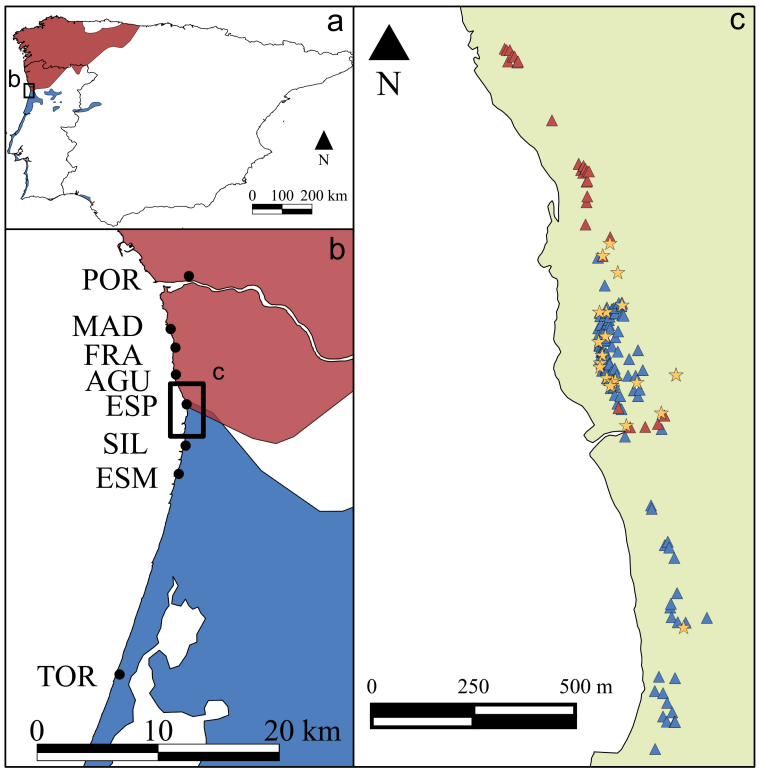

Figure 3.1.1. a) Map of Iberian Peninsula with distribution of *Podarcis bocagei* (in red) and *P. carbonelli* (in blue); **b)** The highlighted area in a) showing the locations of sampled populations in the transect across the hybrid zone; **c)** zoom in of the area in b) where individuals of both parental species were found in syntopy. Each triangle represents an individual assigned to each parental species (Q < 0.1 or Q > 0.9; colours as in a) and b)) and each star represents an individual with *Q* between 0.1 and 0.9 based on Admixture analysis.

DNA extraction and RAD sequencing

Genomic DNA was extracted using the EasySpin® Genomic DNA Tissue Kit (Citomed, Odivelas, Portugal) following the supplier's protocol. The quality and quantity of extracted DNA was evaluated on a 2% agarose gel and a QUBIT 2.0 fluorimeter (Life Technologies, Grand Island, NY, USA).

We obtained RAD sequencing data using modifications to protocols from Parchman *et al.* (2012), Peterson *et al.* (2012) and Purcell *et al.* (2014). The complete protocol is described by Brelsford *et al.* (2016). Briefly, we digested genomic DNA with restriction enzymes *Sbf*I and *Mse*I, ligated barcoded adapters, amplified each individual sample in four independent separate PCR reactions, pooled all PCR products and selected fragments between 400 and 500 bp using a 2.5% agarose gel. The library containing all samples was sequenced on eight Illumina® (San Diego, CA, USA) HiSeq

2000 lanes in the Lausanne Genetic Technology Facility (Lausanne, Switzerland), with single-end 100 bp reads.

Data filtering and SNP calling

We demultiplexed the raw reads using the process_radtags module of Stacks version 1.3 (Catchen 2013) allowing one mismatch for barcodes, to remove low-quality reads, reads with uncalled bases and reads that failed the Illumina 'chastity' filter. Adapters were trimmed from demultiplexed reads using a custom shell script. Reads were then aligned in each individual *de novo* to assemble loci using the denovo_map.pl wrap up program. A *de novo* assembly was chosen over the possibility of an assembly guided by a reference genome since the *P. muralis* genome (which we use in this article) was not yet available at this stage of the work. Each locus was assembled with a minimum depth of sequencing coverage of one and allowing two mismatches between stacks within individuals and two between individuals. We then used VCFtools version 0.1.15 (Danecek *et al.*, 2011) to filter the resulting variants. Loci with depth coverage <8, with missing data >20% or with alleles with minimum frequency <0.05 were removed. We selected one SNP per locus by choosing the SNPs maximizing frequency differences between species with a custom script (available from http://github.com/catpinho).

Population structure across the transect and hybrid identification

Genomic variability among individuals was visualised by performing a principal component analysis (PCA) on the complete dataset using the adegenet package version 2.0.1 (Jombart 2008; Jombart & Ahmed 2011). We used the software Admixture version 1.3.0 (Alexander *et al.* 2009; Alexander & Lange 2011) to evaluate the proportion (Q) of individual genomes from each parental species. Admixture was run until the log-likelihood estimate increases less than 10^{-5} between iterations. We first ran Admixture for K=2 because we were mainly interested in individual admixture between the two species. We also extended these analysis from K=2 to K=6 to detect possible sub-structure within species. The best K was estimated with a five-fold cross-validation error procedure, implemented in Admixture where the best value of K exhibits the lowest cross-validation error. We ran Admixture analyses for the complete dataset and for a dataset consisting of 2300 loci with allele frequencies higher than 0.8 in pure parental

populations of one species and lower than 0.2 in the other ("80/20" dataset).

Simulations of genotypic composition within the contact zone

To investigate the possible causes for the observed distribution of hybrid indexes, we carried out simulations of various scenarios. Particularly we wanted to test if immigration and assortative mating in the contact zone can explain genotypic composition in and if there are post-zygotic barriers to gene flow acting in the contact zone by comparing the proportion of backcrosses regarding the proportions of F1 hybrids. Our scenarios fixed the observed proportion of individuals from the two species (i.e. largely biased towards *P. carbonelli*) and incorporated various levels of immigration of pure genotypes into the contact area and the possibility of a restriction in the proportion of interspecific matings. All simulations started from the same data set of 1000 pure individuals, 182 from one species (mimicking here *P. bocagei*) and 818 from the other (*P. carbonelli*), corresponding to the proportion of both species found in centre of the contact zone. The datasets included data from 1241 loci (the same number of diagnostic loci that we used) all forced to be diagnostic between both species. This initial set up was the same in all scenarios.

In the "random mating" scenarios (RM), there was no restriction to gene flow between any individual in any generation. In the "non random mating" scenarios (NRM) we used the observed proportion of putative F1s (individuals with a hybrid index estimated between 0.4 and 0.6) to "pure" individuals (the total number of individuals with hybrid index lower than 0.1 or higher than 0.9) observed in our data set as the rate of heterospecific matings, ensuring a fixed proportion of F1 to pure individuals in each generation. Apart from this restricted number, no other matings between pure individuals of different species were allowed. There were no restrictions to gene flow both within pure individuals of each species or between hybrids (once they were formed) and any individual (whether "pure" or "hybrid") of the data set. That is, this represents a scenario of "weak" assortative mating in which only matings between pure individuals are restricted.

Immigration from outside of the contact zone was simulated in each generation by randomly replacing a fraction of the individuals by pure genotypes of both species (in the same proportions as they exist in the real data set) resulting in four scenarios of migration (0, 0.2, 0.3 and 0.4 of genotypes replaced). For example, in a scenario of 20%

migration, 20% of the individuals are removed from the data set irrespectively of their genotype and replaced by the same number of individuals with a hybrid index of either 0 or 1 (that is, with a genotype completely characteristic of each species at each of the 1241 loci). For simplicity, the calculation of hybrid indexes inside the simulation procedures was conducted by simply counting the proportion of the "species 1" alleles of each individual. This procedure was repeated up until generation 50, and 100 similar simulations were conducted for each of the tested scenarios. In order to ensure a correct comparison with the data observed in the contact zone, each of the "generation 50" data sets was subjected to a number of transformations: i) 115 individuals - the number collected in the zone of strict syntopy between the species - were chosen at random from the total of 1000 individuals simulated; ii) missing data was added to each of these individuals following the same missing data distribution observed for our data set; iii) genotyping errors were introduced at random up to a maximum value of 0.25% (which we determined to be the maximum genotyping error observed in our data set based on the inspection of replicated individuals); iv) individuals sampled from the extremes of our transect were included so that all data sets have the same reference populations. These data sets were then ran using Admixture in the same exact conditions as the subset of our data including the 115 individuals from the center of the contact zone. Using classes of 10%, we computed the frequency distribution of the individuals' admixture proportions in each simulation. Each of the 100 distributions for each simulated scenario (combination of RM and NRM with 0, 0.05, 0.2, 0.3 and 0.4 of immigrants) was then compared to the real distribution using a series of pairwise Fisher's exact tests and employing the Bonferroni correction for multiple comparisons computed with R package 'rcompanion' version 2.0.0 (Mangiafico 2018). We also compared the average distribution for each scenario, i.e. the average of each class of 10%, to the observed distribution with similar Fisher's exact tests. We then tested for each scenario if the ratios of pure (defined as before) to the total number of individuals (pure/total) and the F1s to backcrosses (all admixed individuals except F1s; F1/BC) were higher that in the observed data. When the proportion of simulations with higher ratios than the observed data was higher than 0.05, we considered that the observed data fit within the distribution generated by the simulations. With these tests we intend to determine if bimodality is stronger than expected in the observed data and if the number of backcrosses is lower in the observed data than expected considering the number of F1s.

All the simulations, format conversions and data set manipulations were performed using a series of scripts available in www.github.com/catpinho.

Geographic cline analyses

To perform analyses of geographic clines we used the R package HZAR (Derryberry *et al.* 2014). This program provides functions to fit allele frequency data to equilibrium geographic cline models (Szymura & Barton 1986, 1991; Barton & Gale 1993; Gay *et al.* 2008) using the Metropolis–Hastings Markov chain Monte Carlo (MCMC) algorithm. For this purpose, for each sampling location was assigned a distance along the transect that corresponded to the shortest distance between it and POR along the coastline (i.e. avoiding a straight line over the ocean). Allele frequencies for each SNP locus included in the "80/20" dataset and the HI estimates were fitted to 15 equilibrium geographic cline models using HZAR version 3.0.3 (Derryberry *et al.* 2014). All models estimated cline centre (distance from POR, c) and width (1/maximum slope, w). Additionally, distinct models could estimate different combinations of the distance from the cline centre to the tail (δ) and the tail slope (τ): no tail (none), right tail only (right), left tail only (left), symmetric tails (mirror), or both tails separately (both); and whether they estimate (free), did not estimate (none) or fixe at 0 and 1 (fixed) the frequencies at cline ends (p_{min} and p_{max}). For this study, right tail means that the model estimates δ and τ to south and left tail to north, i.e. in the direction of *P. carbonelli*, and *P. bocagei*, respectively. We performed three independent runs of 1,000,000 MCMC iterations, a burn-in of 100,000 and sampling every 10 iterations for each model and checked for convergence. For each locus, the model with the lowest AIC score was selected as the best-fitting model. We then estimated the confidence interval (CI) as the region delimited by the maximum and minimum values of the 95% credible cline region (Derryberry et al. 2014). To test if each locus significantly differed from the genome average, we compared the c and w for each locus with the c and w for the hybrid index (HI) cline (c_{HI} and w_{HI}), which we took as a sort of a "genome-wide average". We used the ancestry from the *P. bocagei* cluster as estimated by Admixture (Q) using the complete dataset as a HI. We considered that each parameter for individual clines exhibited significant differences from the HI cline if their 95% CI do not overlap. We also searched for outlier loci on the basis of the geographic location of locus-specific estimates of c and on the size of w. Finally, to test if the number of loci with centres

shifted to north or south were similar or significantly different, we performed a binomial test.

Genomic clines analyses

We quantified patterns of genomic introgression in admixed populations using the BGC software version 1.0 (Gompert & Buerkle 2012). This software uses a Bayesian genomic cline model (Gompert & Buerkle 2011; Gompert *et al.* 2012b; a) to estimate ancestries of individuals summarized as an hybrid index (HI_{BGC}) and to quantify genome-wide variation in introgression among admixed populations. The model includes two parameters (*α* and *β*) that describe the transition in allele frequency for every locus, from the genomic background of origin (ancestry) to the genomic background of the other species, along the gradient of admixed genotype classes (i.e. hybrid index, HI_{BGC}) between the two parental species (Gompert & Buerkle 2011). The genomic cline parameter *α* gives the cline centre, which describes the probability of increased or decreased locus-specific ancestry from one of the species given the HI_{BGC}. In this study, a significantly positive *α* means an increase of *P. bocagei* ancestry relative to the HI_{BGC} and significantly negative *α* means a decrease in the probability of *P. bocagei* ancestry (i.e. an increase of *P. carbonelli* ancestry). The genomic cline parameter *β* describes the rate of transition from low to high probability of *P. bocagei* ancestry as a function of HI_{BGC}, indicating a steeper or shallower change in allele frequencies at a given locus. Significantly positive or negative *β* reveals restricted or increased locus specific introgression, respectively, relatively to the expectations based on the hybrid index.

A priori knowledge of pure parental individuals is required to estimate the proportion of the ancestry attributed to each parental species. We used the populations POR and MAD as *P. bocagei* parental individuals and ESM and TOR as *P. carbonelli*, while using the population ESP as the test (admixed) population. With the complete dataset we ran three independent MCMC chains for 300,000 steps each and recorded samples from the posterior distribution every 25th step following a 200,000 step burn-in to estimate marginal posterior probability distributions for HI_{BGC} and cline parameters *α* and *β*. We combined the output of the two chains after inspecting the MCMC output to assess convergence to the stationary distribution.

The significance of each individual locus *α* and *β* values was assessed on the basis of whether their 95% CI include zero or not. These estimates provide a locus-

specific view on the process of introgression between species but they do not compare how genomic clines differ between loci. Therefore, we additionally searched for outlier loci on the basis of the genome-wide distribution of locus-specific estimates of *α* and *β* values by determining the cline parameter quantiles (Gompert & Buerkle 2011, 2012).

RADtag alignment with Podarcis muralis *genome*

A high-quality, chromosome-level assembly of the *Podarcis muralis* genome (PodMur1.0) with 1.51Gb has been recently made available (Andrade *et al.* submitted). *P. muralis* is a close relative of the Iberian and North African clade of *Podarcis* and thus this is a good reference genome for this study. We aligned each of the 6905 RADtag sequences from the complete SNP dataset with the *P. muralis* genome using the blastn algorithm included in the BLAST package (Altschul *et al.* 1997) and observed whether loci with similar patterns of introgression were clumped together in large regions putatively corresponding to genomic islands. We further aimed to examine if the Z chromosome would have a major effect on reproductive isolation than autosome chromosomes since several lines of evidence suggest that both Z (Sætre *et al.* 2003; Kirkpatrick & Hall 2004; Albert & Otto 2005; Hall & Kirkpatrick 2006; Storchová *et al.* 2010) and X chromosomes (Orr 1987; Tucker *et al.* 1992; Storchova *et al.* 2004) play a key role in reproductive isolation. We tested if the number of loci with a specific kind of introgression pattern in Z chromosome was distinct from those in the autossomes by performing a Fisher's exact test with R package 'rcompanion' version 2.0.0 (Mangiafico 2018) for each of the geographic and genomic significant sets of loci. We further performed a one-way ANOVA between Z chromosome and the autosomes for each geographic and genomic cline parameter to test if there was any difference between the parameter values of the Z chromosome and the autosomes.

Results

Data filtering and SNP calling

The final dataset, after removing loci with depth coverage <8, missing data >20% and individuals with more than 30% of missing data, consisted of 6905 SNPs (complete dataset) with an average depth of coverage of 27.6. We further filtered out individuals with more than 30% of missing data, resulting in a final dataset with 329 individuals with an average depth of coverage of 27.6, too. The replicability was higher than 99% for all

samples. Additionally, we filtered the complete dataset to obtain a dataset consisting of loci with allele frequencies higher than 0.8 in pure parental populations of one species and lower than 0.2 in the other, resulting in a dataset of 2300 loci ("80/20" dataset).

Population structure across the transect

As expected from previous works, the current hybrid zone between *P. bocagei* and *P. carbonelli* is restricted to the ESP population and mostly consists in pure individuals from both species (Figures 3.1.2. and 3.1.3.). Actual syntopy was restricted to a ~450 meters wide area in ESP where we collected 104 pure parental individuals of both species, 13 identified as *P. bocagei* and 91 as *P. carbonelli* based on multilocus assignment results for K=2. Binomial probability (p<0.001) shows that the number of samples in this area is biased towards *P. carbonelli*, indicating a higher density of this species since nothing suggests higher capturability. Eleven individuals of mixed ancestry ($0.1 < Q < 0.9$) were identified in ESP for K=2 inside the area of syntopy but two were also found outside the syntopy area in ESP. Accordingly, the first PCA axis (PC1) explains 57.1% of the variation in the data and clearly distinguishes *P. bocagei* and *P. carbonelli* populations (Figure 3.1.2). PC2 explains 1.4% of the variation and corresponds to the genetic structure within *P. carbonelli*, separating the ESP population from all other *P. carbonelli* populations. In fact, when increasing K with Admixture, K=3 is the best scenario to explain the variation in the SNP data (Appendix II, Figure S2.1.). In concordance with the PCA, *P. carbonelli* individuals from ESP are mostly assigned to a group apart from the other *P. carbonelli* populations showing sub-structuring within *P. carbonelli*.

An admixture signal is also evident in SIL population south to ESP since all individuals in SIL were identified as *P. carbonelli* on the basis of morphology but 0.02 to 0.05 of admixture from *P. bocagei* was evident from SNP data (Figure 3.1.3.), a higher proportion than in any other *P. carbonelli* population away from the hybrid zone.

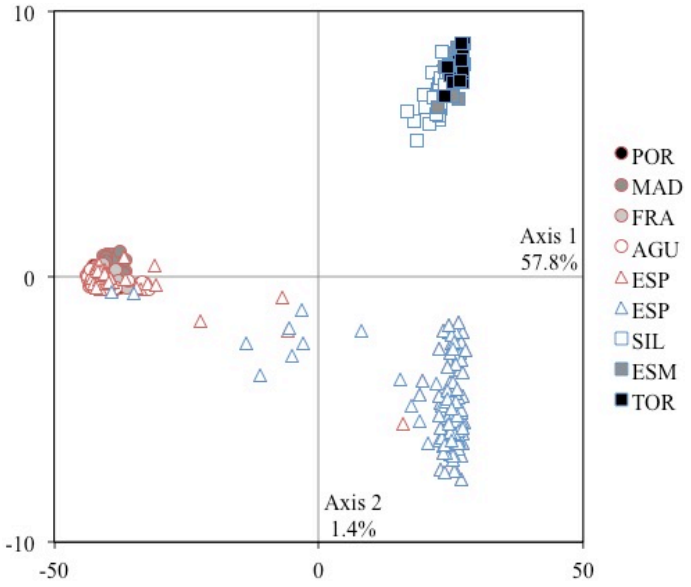

Figure 3.1.2. Principal Component Analysis (PCA) of 6905 SNP loci in the 329 individuals calculated with Adegenet R package. Circles represent individuals from populations north of the contact zone, triangles correspond to the individuals from the contact zone and squares identify the individuals from populations south of the contact zone. Red represents individuals identified as *P. bocagei* and blue as *P. carbonelli* based on field identification. Population acronyms as in Figure 3.3.1.

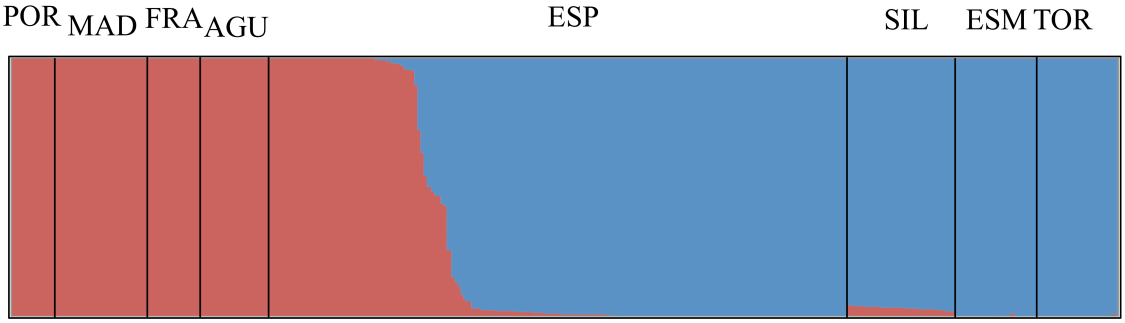

Figure 3.1.3. Results from individual multilocus genotype clustering computed with Admixture for the 6905 loci. Each individual is represented by a vertical line proportionally partitioned into the K=2 coloured segments. Population acronyms and colours as in Figure 3.1.1.

Simulations of genotypic composition within the contact zone

Simulations suggested that most combinations of random or assortative mating with various migration rates are unable to produce distributions of genotype classes matching our observed distribution (Figure 3.1.4.). Only for non-random mating with a rather unrealistic 40% of parental genotypes migrating into the contact zone, the simulations are not significantly different from our data (*p*-value = 0.052 for the comparison of average simulated scenario with the observed data, significance level = 6.25×10^{-3} after Bonferroni correction; 91 simulations do not significantly differ from the observed data; Appendix II, Table S2.1.). Without such unrealistic migration rates, F1 hybrid genotypes disappear in a few generations both with random and non-random mating.

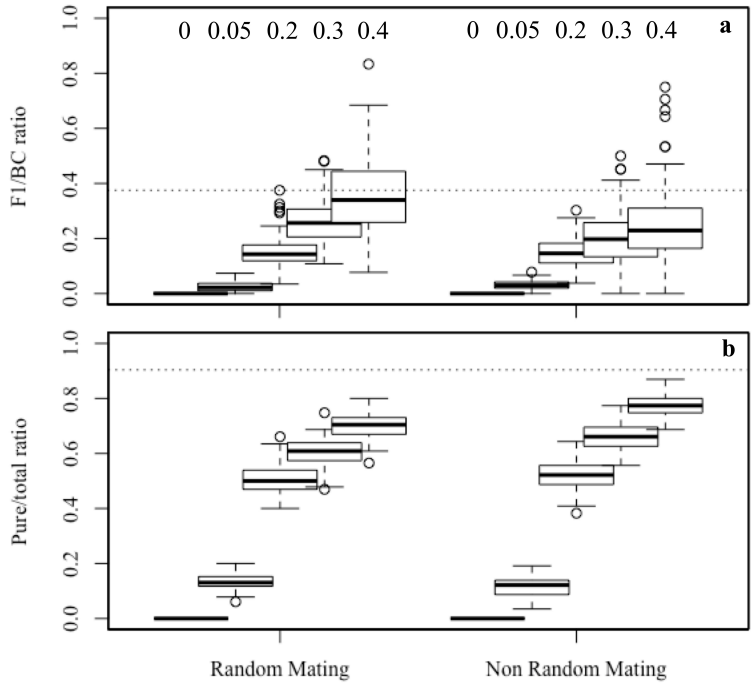

Figure 3.1.4. Boxplots representing **a)** the simulated distributions of F1 hybrids by backcrosses ration (F1/BC) and **b)** pure parental by total ratio (Pure/total) for each scenario combining random and non-random mating with several migration rates. The numbers on top represent each migration rate aligned with the respective boxplots. Horizontal doted lines represent the same ratios in the observed data.

For most scenarios with random and non-random mating, both proportions of F1 relative to backcrosses and pure parentals relative to total individuals were significantly higher in observed data (Appendix II, Table S2.2.). This suggests less backcrosses than expected in the observed data given the number of F1 hybrids and that the bimodality in the observed data is higher than in any of the simulated scenarios, respectively.

Geographic and genomic clines

Geographic cline analysis revealed that the model with right tail fitted (i.e. estimated tail in the direction of *P. carbonelli*) and p_{min}/p_{max} not estimated (right/none) was the best-fitted for the HI cline, with centre c_{HI} = 13150m (CI = 11980–14040m), corresponding to 13 km south of POR (ESP; Figure 3.1.5.) and width w_{HI} = 540m (CI = 100–3180m) indicating a very narrow hybrid zone. Most of the geographic clines for each SNP locus were highly concordant in both *c* (median = 13850m, mean =13420m, SD = 1250m, range = 1910–18960m; Table 3.1.1. and Figure 3.1.5.) and *w* (median = 890m, mean = 1850m, SD = 2630m, range = 10–26010m; Table 3.1.1. and Figure 3.1.5.) and not significantly different from the HI estimates. The best-fitted models varied across loci (see Appendix II, Table S2.3.): the most common were the models where no tails were fitted and p_{min}/p_{max} were estimated (none/free; 583 loci), only right tail fitted

and no p_{min}/p_{max} estimated (right/none; 438 loci) and no tail fitted and without p_{min}/p_{max} estimated (none/none; 415 loci). Only 120 loci (5.2%) exhibit cline centres that differ significantly from the genomic average (as estimated by HI): 73 loci (3.2%) have a cline that is shifted north (range = 1900–11270m south of POR) and 47 (2%) shifted south (range = 14160m–19150m south of POR). The binomial test revealed that these proportions are significantly different (p<0.001) than random expectations. Moreover we only detected outliers for *c* shifted to north in the six loci with *c* closer to POR population (range = 1900m–4350m south of POR). One hundred and eighty-seven loci (8.1%) have significantly wider clines (min = 4020m; max = 26010m) than the genomic average but none has a narrower cline. Detecting statistical significant lower *w* may not be possible since w_{HI} is already very low and so is the lower margin of the confidence interval (w_{HI} = 540m, CI = 100–3180m). The regression model estimated for *c* versus *w* suggests that there is no significant linear correlation between these two parameters, although most clines with *c* shifted to north tend to have larger *w*.

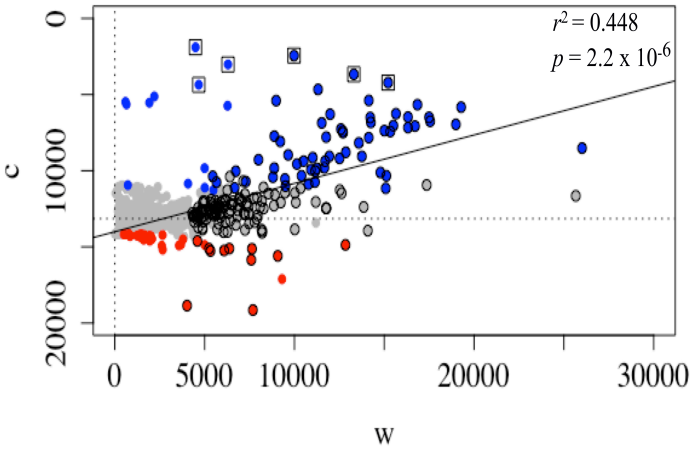

Figure 3.1.5. Cline variation of 2300 loci ("80/20" dataset) described by *c* and *w* estimated with HZAR. Each point represents each locus. Coloured points represent loci with *c* significantly shifted to south regarding c_{HI} (red) and to north (blue). Loci with black outlines have *w* significantly higher than w_{HI}. Squares denote loci with outlier *c*. Horizontal doted line represent the c_{HI} (13150m) and the vertical represents the w_{HI} (540m). The solid line represents the regression line between *c* and *w* values.

In genomic clines we observed more heterogeneity across the genome (-6.95<α<5.15; -18.31<β<29.90) with many loci significantly deviating from the expectations given the HI estimated from total SNP datasets (Table 3.1.1. and Figure 3.1.6.). We detected a significant excess of *P. bocagei* ancestry (lower bound of 95% CI for $\alpha > 0$) for 1484 loci (21.5%) and an excess of *P. carbonelli* ancestry (upper bound of 95% CI for $\alpha < 0$) for 2223 loci (32.2%). These proportions are significantly different than expected by chance (p<0.001). A total of 3198 loci (46.3%) do not present an excess of ancestry for any species given the HI (95% CI for α includes 0). We also detected 1750

loci (25.3%) with significantly steeper changes in allele frequencies from one species to another (lower bound of 95% CI for $\beta > 0$), 1473 loci (21.3%) with significantly shallower changes in allele frequencies (upper bound of 95% CI for $\beta < 0$) and 3682 loci (53.3%) following the expectations given the HI (95% CI for β includes 0). From loci with significantly positive α, 237 (3.4% of the total) have β not significantly different from zero, 303 (4.4% of the total) have significantly positive β and 944 (13.7% of the total) show significantly negative β. For loci with significantly negative α, β does not significantly differ from zero in 1051 (15.2% of the total), 1092 (15.8% of the total) loci show significantly positive β and only 80 (1.1% of the total) have significantly negative β. These results suggest an apparent association in some loci. Most loci with increased *P. bocagei* ancestry also have increased introgression while about half of the loci with increased *P. carbonelli* ancestry have restricted introgression. No outlier loci were detected given the distributions of α and β.

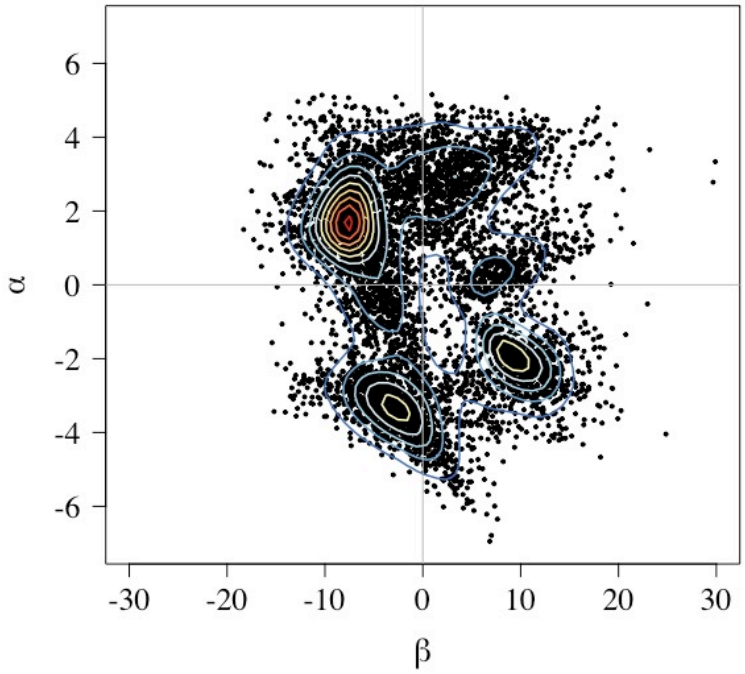

Figure 3.1.6. Heterogeneity of introgression in the contact zone described by α and β. Points represent each 6905 loci (complete dataset) and contour lines show regions of the $\alpha-\beta$ parameter space with similar density of loci.

By comparing geographic and genomic clines we did not detect any correlation between *c* and α or between *w and* β (Figure 3.1.7.). Most loci with significant (positive or negative) α have their *c* close to c_{HI} (13.15 km). The main deviation to this pattern is that 43 loci with *c* shifted to north, out of the 73, have significant negative α (*P. carbonelli* ancestry). By comparing *w and* β we understand that generally β has a larger variation

than *w*. Most loci with significant (positive or negative) *β* have a low *w*. While for positive *β* such results seem concordant (restricted introgression), for negative *β* (increased introgression) these results contrast with the small *w* (restricted introgression).

Table 3.1.1 Summary results of geographic and genomic cline analysis performed respectively with HZAR on 2300 loci ("80/20" dataset, upper table) and BGC on 6905 loci (complete dataset, lower table). Shaded columns represent the number of loci (N) falling in each class of the cline centre (*c*)/cline width (*w*) and locus specific ancestry (*α*)/introgression rate (*β*), from geographic and genomic cline analysis, respectively.

		w							
		Similar to HI		Smaller than HI		Higher than HI		Total	
		N	%	N	%	N	%	N	%
c	Not shifted	2064	89.74	0	0.00	116	5.04	2180	94.87
	Shifted north	13	0.56	0	0.00	60	2.61	73	3.17
	Shifted south	36	1.57	0	0.00	11	0.48	47	2.04
	Total	3113	91.87	0	0.00	187	8.13	2300	100

		β							
		Zero		Negative		Positive		Total	
		N	%	N	%	N	%	N	%
α	Zero	2394	34.67	449	6.50	355	5.14	3198	46.31
	Positive	237	3.43	944	13.67	303	4.39	1484	21.49
	Negative	1051	15.22	80	1.16	1092	15.81	2223	32.19
	Total	3682	53.32	1473	21.33	1750	25.34	6905	100

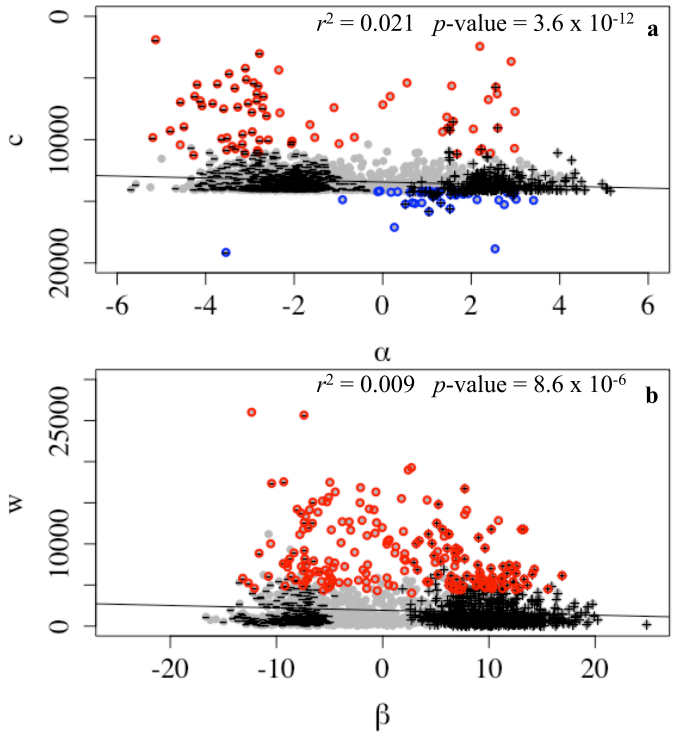

Figure 3.1.7. a) Concordance of *α* and *c*. Grey points represent each 2300 loci common to both complete and "80/20" dataset. Coloured circles represent loci with *c* shifted from c_{HI} to north (red) or to south (blue). Loci with *α* significantly different from zero are represented with "-" (negative) and "+" (positive). **b)** Concordance of *w* and *β*. Circles represent each 2300 loci common to both complete and "80/20" dataset. Coloured circles represent loci with increased (blue) or decreased (red) *β* relatively to neutral expectations. Loci with *w* significantly higher from w_{HI} are represented with "+" (loci with w significantly lower from w_{HI} were not detected). In both plots, the solid line represents the regression line between *c* and *α* between *w* and *β* values.

RADtag alignment with Podarcis muralis *genome*

We used the 6905 RADtag sequences containing the SNPs included in the complete dataset to identify their location in the 1.51 Gb scaffold genome of *P. muralis* genome. We successfully aligned 6027 RADtags with identified regions of the genome (Figure 3.1.8.). This represents a density of 1 marker per 230Kb. Focusing only on the Z chromosome, the density increases to 1.7 SNPs per 230Kb. Considering the sets of loci with exceptional genomic cline parameter estimates, we found that the Z chromosome is enriched for loci with significant positive *α* (i.e. significant excess of *P. bocagei* ancestry) comparing to autosomes while the other sets of significant genomic clines parameters are lower (Appendix II, Table S2.4.). The proportions of loci with significant geographic cline parameters do not present any significant difference (Appendix II, Table S2.4.). ANOVA results reveal that the mean estimated values of all parameters were significantly different between the sexual chromosome and the autosomes (Figure 3.1.9.). Moreover the differences detected between the two genomic regions were highly concordant between geographic and genomic clines. Mean *c* from Z chromosome loci is higher than mean *c* from autosomes, i.e. more shifted to south, and mean *α* from Z chromosome is negative while in autosomes is close to zero, i.e. higher *P. carbonelli* ancestry. The mean *w* of geographic clines of loci from Z chromosome is narrower than from autosomes and the mean *β* in Z chromosome is positive (restricted introgression) while in autosomes is zero.

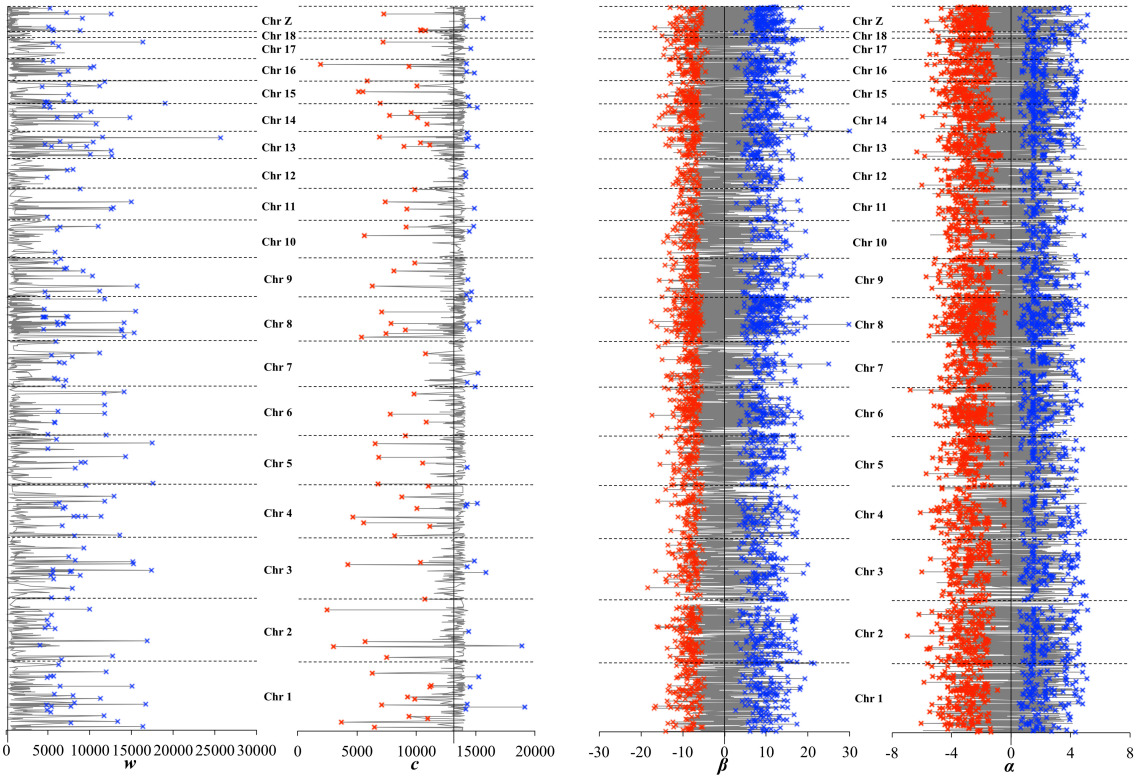

Figure 3.1.8. Variation of geographic cline centre (*c*) and width (*w*) of the 2300 loci ("80/20" dataset) estimated with HZAR across the genome (left panel) and variation of genomic cline centre (*α*) and introgression rate (*β*) of the 6905 loci (complete dataset) calculated with BGC (right panel). Vertical axis represents the 1.47Gb genome size. Each section (Chr 1 to 18) delimited by horizontal dashed lines represents each chromosome and is proportional to its size. The underlined Chr Z represents the sexual chromosome. Grey lines represent the variation of each parameter across the genome. Each coloured "x" represents loci with parameters that significantly deviate from genomic average (red for loci significantly lower and blue for loci significantly higher). Vertical axis crosses horizontal axis at genomic average.

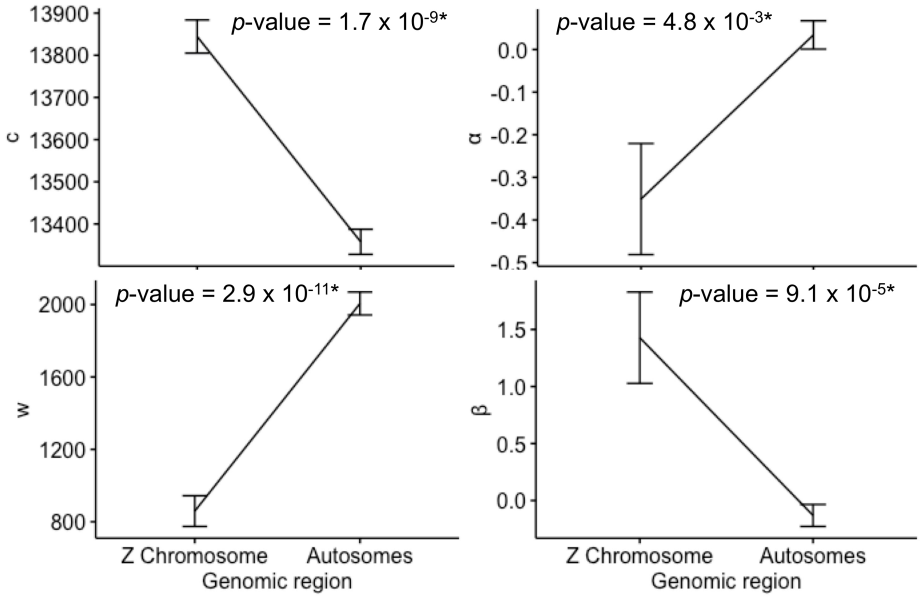

Figure 3.1.9. Mean estimated values in each set of geographic and genomic cline parameters for the Z chromosome and autosomes. ANOVA *p*-values are represented for each set. * represents significant different means.

Discussion

We performed a detailed study on patterns of introgression between *P. bocagei* and *P. carbonelli* in the only natural contact zone between the two species. Our results demonstrate that there are important intrinsic and extrinsic barriers against interspecific gene flow despite incomplete reproductive isolation. We found that regions potentially involved in reproductive isolation are not grouped in few regions but are rather spread over the genome, potentially under distinct selective forces. The role of Z chromosome in reproductive isolation seems to be distinct from that of the autosomes. The genetic basis of reproductive isolation in late stages of speciation is complex and can be explained by multiple, complementary explanations that are discussed below.

Population structure

As shown by previous works, the hybrid zone between *P. bocagei* and *P. carbonelli* is very narrow and mostly built by parental genotypes. Along the 450 meters where individuals of both species occur in syntopy we found a small proportion of admixed individuals (12.42%) compared to the large parental proportions, particularly *P. carbonelli*. Among admixed individuals, the heterogeneity of assignment proportions suggest that hybrids do backcross with parental individuals.

Within *P. carbonelli* we detected genetic structure. The individuals from ESP exhibit a genetic composition different from other *P. carbonelli* populations across the transect, which may be attributed to divergence after restricted gene flow between ESP and southern populations. Alternatively, migration from other unsampled populations (e.g. inland populations) is also a plausible explanation for this pattern.

Strong extrinsic and intrinsic barriers against gene flow

The comparison of the observed data with simulated distributions of admixture classes within the hybrid zone reveals two main incongruences. First, most of the simulated distributions do not match the data observed in the syntopy area. The simulated scenarios were similar to the observed data only when we considered highly unrealistic migration levels (i.e. the replacement of 40% of the individuals in the contact zone by migrants in each generation). Such results indicate that only dispersal and the simple scenario of assortative mating simulated here cannot explain the highly bimodal distribution of admixture classes in the observed data. The observed pattern may thus

be explained by the effects of selection acting on hybrids and/or the presence of other stronger barriers to interspecific gene flow. Second the ratios pure/total and F1/BC are always higher in the observed data. The higher ratio pure/total is due to a higher number of pure individuals in the observed data, i.e. due to a higher bimodality. Moreover, the higher ration F1/BC suggests that the frequency of backcrosses is lower in the observed data than it would be expected given the number of F1s. This can be explained by the existence of barriers to gene flow that maintain a high bimodality in this hybrid zone.

The centre of the large majority of geographic clines (94.9%) was coincident with the area where the two species co-occur. The estimated HI cline indicates that the centre of the hybrid zone is located about 13150m south of POR population. Also most geographic clines (91.9%) exhibit narrow widths that do not differ significantly from the HI cline width, which is around 540m. Steep clines indicate the presence of strong barriers to gene flow that avoid extensive introgression between populations apart from the contact zone (Mallet *et al.* 1990). Only a small fraction of the loci contrast with this pattern. The proportion of loci that deviate to north (3.2%) is statistically different than the proportion deviated to south (2%). While cline' centres of loci shifted to south are all deviated few hundred meters next to the area of syntopy contact, the centre of clines shifted to north range from the area between POR to AGU. Among the former group of loci, 8% were detected as outliers and 82% are wider than the HI cline. Such results suggest that these regions are good candidates for harbouring genes with *P. carbonelli* alleles under positive selection in the genomic background of *P. bocagei*. Loci with cline centres shifted from the majority of other clines has been interpreted as subject to positive selection on one side of the hybrid zone (Barton & Hewitt 1985; Teeter & Payseur 2008; Baldassarre *et al.* 2014).

In contrast to the limited introgression outside the contact zone and the overall concordance among loci, we found large heterogeneity of introgression patterns within the contact zone among SNPs as revealed by genomic cline analyses. Here the lack of outlier loci may be explained due to the lack of concordance for the large majority of loci, which do not allow detecting loci with outlier behaviour. About 25% of the loci exhibit decreased patterns of introgression (significantly positive β) suggesting strong barriers to interspecific gene flow for these genomic regions. This is roughly equivalent to the proportion of loci showing increased introgression (about 21% of loci with significantly negative β). In earlier stages of speciation it is expected that most of these regions with

increased introgression would flow freely between diverging populations (Payseur 2010). This is not the case, however, of *P. bocagei* and *P. carbonelli* as demonstrated before with geographic cline results, resembling an observed pattern of introgression at late stages of speciation, but where only a subset of the genome confers reproductive barriers.

Commonly epistatic interactions (DM incompatibilities) are pointed as the most likely path to the evolution of intrinsic barriers to gene flow (Coyne & Orr 2004; Qvarnström & Bailey 2009) and are expected to generate patterns of differential introgression because of stochasticity in the accumulation of such incompatibilities (Fishman *et al.* 2001; Scascitelli *et al.* 2010; Brennan *et al.* 2014). However, distinct selective forces may have similar consequences on patterns of introgression. Selection against hybrid genotypes, whether arising from strong selection against single locus interspecific heterozygotes (underdominance) in the presence of high levels of gene flow form parental populations or assortative mating, prompts also the increase of β (i.e. restricted introgression) because alleles will not cross species boundaries or alleles inherited from each parental species will be confined to F1 hybrids that do not backcross (Gompert *et al.* 2012b). This effect may be also pronounced in the presence of population structure within the hybrid zone (Gompert & Buerkle 2011; Gompert *et al.* 2012b). The high positive estimates of β most likely arose due to DM incompatibilities and assortative mating but in some loci they may be attributed to population structure within *P. carbonelli*. Furthermore, loci with estimates of α significantly differing from zero (i.e. genomic cline centres deviated towards one or the other genomic background) can also arise from several forms of selection, like underdominance (Gompert & Buerkle 2011; Gompert *et al.* 2012b; Jones *et al.* 2013), DM incompatibilities or directional selection (Gompert & Buerkle 2011).

Notably, we observed an asymmetry in restrictions to introgression in loci with shifted genomic cline centres. Among loci with increased *P. bocagei* ancestry (significant positive α), the proportion of loci with increased introgression (significant negative β; 13.7%) is considerably higher than the proportion of loci with restricted introgression (significant positive β; 4.4%), suggesting that several loci with excess *P. bocagei* ancestry can introgress in the genomic background of *P. carbonelli*. On the other hand the proportion of loci with increased *P. carbonelli* ancestry (significant negative α) and with increased introgression (significant negative β; 1.2%) is much lower than those with

restricted introgression (positive non-zero *β*; 15.8%), suggesting that many loci with increased *P. carbonelli* ancestry have barriers to introgression in the genomic background of *P. bocagei*. Such results indicate directional introgression within the hybrid zone potentially mediated by asymmetric intrinsic barriers to gene flow.

A comparison of geographic and genomic cline parameters failed to detect any correlation. Concordant steep geographic clines contrasts with more heterogenous pattern of introgression among loci within the hybrid zone. This observation evidences the existence of extrinsic barriers to gene flow outside the contact zone for the large majority of the loci. These two species contact in the extremes of their distributions in a very narrow stripe along the coastline. This might be explained by the fact that both species present overall distinct climatic niches, despite a partial overlap (Caeiro-Dias *et al.* 2018), due to their preference for habitats with relatively high humidity (Caeiro-Dias *et al.* 2018) and consequent association with Atlantic climatic regimes (Sillero *et al.* 2009). This area corresponds also to the distribution limits of other species, including some *Podarcis* species (e.g. *Pelodytes punctatus, Podarcis guadarramae lusitanicus*, *Podarcis virescens*; Loureiro *et al.* 2008). We thus speculate here that climatic factors, most likely together with other environmental factors that remain to be identified, may be acting as extrinsic barriers to gene flow between *P. bocagei* and *P. carbonelli*, preventing introgression outside the contact zone. Consequently, reproductive isolation is not maintained only by intrinsic features but also by interactions with the environment.

Reproductive isolation across the genome

Distinct kinds of intrinsic and extrinsic selective forces are contributing to the differential pattern of introgression observed within a contact zone and to the contrasting introgression patterns within and outside the contact zone between two species in a late stage of speciation. When mapping on *P. muralis* genome, we found that loci exhibiting increased or decreased patterns of introgression are distributed across the genome, suggesting that genomic regions potentially involved in speciation are not clumped in particular genomic regions. Similar patterns were also observed in the initial stages of speciation (Nakazato *et al.* 2007; Michel *et al.* 2010) and are expected if adaptive traits or barriers to gene flow are caused by several widespread alleles with small individual effects (Parchman *et al.* 2013). It can arise in divergent allopatric populations, as genetic divergence may accumulate virtually anywhere in the genome, if selection acts on

standing genetic variation, favouring many alleles with weak advantage (Hermisson & Pennings 2005; Pritchard *et al.* 2010). Also, patterns of allopatric divergence are likely the best explanation for climatic niche divergence across *P. hispanicus* complex (Caeiro-Dias *et al.* 2018) and thus the patterns of reproductive isolation between *P. bocagei* and *P. carbonelli* may be related with divergence in allopatry, where a large portion of the genome diverged through diverse evolutionary processes besides divergent selection (Gompert *et al.* 2012a). For example, epistatic gene interactions can arise from neutral processes (Gavrilets *et al.* 1998; Lynch & Force 2000; Fierst & Hansen 2010).

When we compared the patterns of introgression between the Z chromosome and the autosomes, we discovered that the proportion of loci with exceptional geographic clines relative to the HI seem to be similar between the Z chromosome and the autosomes. However, we found that in the Z chromosome the loci with significant genomic cline parameters have distinct patterns than in autosomes. This last observation is expected when the sexual chromosomes have a distinct role in the reproductive isolation compared to autosomes. Many studies have identified a distinct role of the sexual chromosomes (Z chromosome or X chromosome) in introgression patterns and reproductive isolation (Orr 1987; e.g. Guénet *et al.* 1990; Jiggins *et al.* 2001; Storchová *et al.* 2010; Llopart 2012), which is a long-standing principle of speciation – the “large X effect”. We found that loci with restricted introgression detected with genomic clines were proportionally less in Z chromosome than across autosomes but on average such loci introgress less. These results suggest that intrinsic genomic incompatibilities may be less common in Z chromosome than in autosomes but may have a higher influence on reproductive isolation. For instance, if alleles are recessive, hemizygous X (or Z) chromosome substitutions have a much larger effect than autosomal substitutions (Coyne & Orr 2004).

Is the hybrid zone between P. bocagei *and* P. carbonelli *moving?*

Interestingly, in SIL, the population immediately south of the contact zone, all the individuals were identified as *P. carbonelli* in the basis of morphology, but a small proportion of the assignment with SNP data (0.02 to 0.05) was attributed to *P. bocagei*, which was interpreted as an admixture signal rather than an analytical artefact, since for K=3 (best K) the admixture signal still persists. Additionally, in the PCA some samples from SIL are slightly deviated towards other intermediate individuals between *P. bocagei*

and *P. carbonelli*. The consistency in the proportion of assignment to *P. bocagei* across SIL individuals suggests that this population suffered the effect of admixture in the past, with subsequent restriction in the gene flow leading to the homogenization of current assignment proportions. Also, a large fraction of steep clines had a steep tail to north and a smother tail to south. Overall, such results seem to be consistent with a scenario of a moving hybrid zone northwards in the direction of *P. bocagei*, since a moving hybrid zone is expected to leave a tail of clines of unlinked neutral markers resulting in an asymmetrical cline variation (Buggs 2007; Teeter *et al.* 2010).

Conclusion remarks

In this study we found consistent patterns of overall strong reproductive isolation between two deep divergent *Podarcis* species, which is consistent with late stages of speciation. We found evidence for the existence of both intrinsic and extrinsic barriers to extensive interspecific gene flow for the large majority of the loci. Loci potentially involved in intrinsic barriers are well distributed across the genome, suggesting that several regions across the genome under distinct selective forces contribute to reproductive isolation in the late stages. Moreover, the Z chromosome was found to have a distinct role in the reproductive isolation relatively to the autosomes. Particularly, the introgression rate in the Z chromosome seems to be lower that in autosomal chromosomes. The proportion of assignment to *P. bocagei* across SIL individuals and the asymmetrical cline variation in some loci seems consistent with a moving hybrid zone scenario. Overall, the results reported here denote that the genomic architecture of reproductive isolation in the late stages of speciation relies on cohesion across the genome against interspecific gene flow and is influenced by complex combination of multiple interacting barriers.

References

Albert AYK, Otto SP (2005) Sexual selection can resolve sex-linked sexual antagonism. *Science*, **310**, 119–121.

Alexander DH, Lange K (2011) Enhancements to the ADMIXTURE algorithm for individual ancestry estimation. *BMC Bioinformatics*, **12**, 246.

Alexander DH, Novembre J, Lange K (2009) Fast model-based estimation of ancestry in

unrelated individuals. *Genome Research*, **19**, 1655–1664.

Altschul SF, Madden TL, Schäffer AA *et al.* (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research*, **25**, 3389–3402.

Andrew RL, Rieseberg LH (2013) Divergence is focused on few genomic regions early in speciation: incipient speciation of sunflower ecotypes. *Evolution*, **67**, 2468–2482.

Baldassarre DT, White TA, Karubian J, Webster MS (2014) Genomic and morphological analysis of a semipermeable avian hybrid zone suggests asymmetrical introgression of a sexual signal. *Evolution*, **68**, 2644–2657.

Barton NH, Gale KS (1993) Genetic analysis of hybrid zones. *Hybrid zones and the evolutionary process*, 13–45.

Barton NH, Hewitt GM (1985) Analysis of hybrid zones. *Annual Review of Ecology and Systematics*, **16**, 113–148.

Barton NH, Hewitt GM (1989) Adaptation, speciation and hybrid zones. *Nature*, **341**, 497.

Brelsford A, Dufresnes C, Perrin N (2016) High-density sex-specific linkage maps of a European tree frog (*Hyla arborea*) identify the sex chromosome without information on offspring sex. *Heredity*, **116**, 177.

Brennan AC, Hiscock SJ, Abbott RJ (2014) Interspecific crossing and genetic mapping reveal intrinsic genomic incompatibility between two *Senecio* species that form a hybrid zone on Mount Etna, Sicily. *Heredity*, **113**, 195.

Buggs RJA (2007) Empirical study of hybrid zone movement. *Heredity*, **99**, 301.

Caeiro-Dias G, Luís C, Pinho C, Crochet PA, Sillero N, Kaliontzopoulou A (2018) Lack of congruence of genetic and niche divergence in *Podarcis* hispanicus complex. *Journal of Zoological Systematics and Evolutionary Research*, **56**, 479–492.

Catchen JM (2013) Stacks: an analysis tool set for population genomics. *Molecular Ecology*, **22**, 3124–3140.

Coyne JA, Orr HA (2004) Speciation. Sunderland, MA.

Cruickshank TE, Hahn MW (2014) Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. *Molecular Ecology*, **23**, 3133–3157.

Derryberry EP, Derryberry GE, Maley JM, Brumfield RT (2014) HZAR: hybrid zone analysis using an R software package. *Molecular Ecology Resources*, **14**, 652–663.

Dobzhansky TG (1937) *Genetics and the Origin of Species*. Columbia university press.

Ellegren H, Smeds L, Burri R *et al.* (2012) The genomic landscape of species divergence in *Ficedula* flycatchers. *Nature*, **491**, 756.

Feder JL, Egan SP, Nosil P (2012a) The genomics of speciation-with-gene-flow. *Trends in Genetics*, **28**, 342–350.

Feder JL, Gejji R, Yeaman S, Nosil P (2012b) Establishment of new mutations under divergence and genome hitchhiking. *Philosophical Transactions of the Royal Society B*, **367**, 461–474.

Feder JL, Nosil P (2010) The efficacy of divergence hitchhiking in generating genomic islands during ecological speciation. *Evolution*, **64**, 1729–1747.

Fierst JL, Hansen TF (2010) Genetic architecture and postzygotic reproductive isolation: evolution of Bateson–Dobzhansky–Muller incompatibilities in a polygenic model. *Evolution*, **64**, 675–693.

Fishman L, Kelly AJ, Morgan E, Willis JH (2001) A genetic map in the *Mimulus guttatus* species complex reveals transmission ratio distortion due to heterospecific interactions. *Genetics*, **159**, 1701–1716.

Fitzpatrick BM (2013) Alternative forms for genomic clines. *Ecology and Evolution*, **3**, 1951–1966.

Gavrilets S, Hai L, Vose MD (1998) Rapid parapatric speciation on holey adaptive landscapes. *Proceedings of the Royal Society of London B: Biological Sciences*, **265**, 1483–1489.

Gay L, Crochet PA, Bell DA, Lenormand T (2008) Comparing clines on molecular and phenotypic traits in hybrid zones: A window on tension zone models. *Evolution*, **62**, 2789–2806.

Gompert Z, Buerkle CA (2009) A powerful regression-based method for admixture mapping of isolation across the genome of hybrids. *Molecular Ecology*, **18**, 1207–1224.

Gompert Z, Buerkle CA (2011) Bayesian estimation of genomic clines. *Molecular Ecology*, **20**, 2111–2127.

Gompert Z, Buerkle CA (2012) bgc: Software for Bayesian estimation of genomic clines. *Molecular Ecology Resources*, **12**, 1168–1176.

Gompert Z, Forister ML, Fordyce JA *et al.* (2010) Bayesian analysis of molecular variance in pyrosequences quantifies population genetic structure across the

genome of *Lycaeides* butterflies. *Molecular Ecology*, **19**, 2455–2473.

Gompert Z, Lucas LK, Nice CC *et al.* (2012a) Genomic regions with a history of divergent selection affect fitness of hybrids between two butterfly species. *Evolution*, **66**, 2167–2181.

Gompert Z, Mandeville EG, Buerkle CA (2017) Analysis of population genomic data from hybrid zones. *Annual Review of Ecology, Evolution, and Systematics*, **48**, 207–229.

Gompert Z, Parchman TL, Buerkle CA (2012b) Genomics of isolation in hybrids. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **367**, 439–450.

Gourbiere S, Mallet J (2010) Are species real? The shape of the species boundary with exponential failure, reinforcement, and the "missing snowball." *Evolution*, **64**, 1–24.

Guénet J-L, Nagamine C, Simon-Chazottes D, Montagutelli X, Bonhomme F (1990) Hst-3: an X-linked hybrid sterility gene. *Genetics Research*, **56**, 163–165.

Hall DW, Kirkpatrick M (2006) Reinforcement and sex linkage. *Evolution*, **60**, 908–921.

Harrison RG, Larson EL (2016) Heterogeneous genome divergence, differential introgression, and the origin and structure of hybrid zones. *Molecular Ecology*, **25**, 2454–2466.

Hermisson J, Pennings PS (2005) Soft sweeps: molecular population genetics of adaptation from standing genetic variation. *Genetics*, **169**, 2335–2352.

Iyengar VK, Reeve HK, Eisner T (2002) Paternal inheritance of a female moth's mating preference. *Nature*, **419**, 830.

Janoušek V, Wang L, Luzynski K *et al.* (2012) Genome-wide architecture of reproductive isolation in a naturally occurring hybrid zone between *Mus musculus musculus* and *M. m. domesticus*. *Molecular Ecology*, **21**, 3032–3047.

Jiggins CD, Linares M, Naisbit RE *et al.* (2001) Sex-linked hybrid sterility in a butterfly. *Evolution*, **55**, 1631–1638.

Jombart T (2008) adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics*, **24**, 1403–1405.

Jombart T, Ahmed I (2011) adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. *Bioinformatics*, **27**, 3070–3071.

Jones MR, Forester BR, Teufel AI *et al.* (2013) Integrating landscape genomics and spatially explicit approaches to detect loci under selection in clinal populations. *Evolution*, **67**, 3455–3468.

Kaliontzopoulou A, Pinho C, Harris DJ, Carretero MA (2011) When cryptic diversity blurs the picture: A cautionary tale from Iberian and North African *Podarcis* wall lizards. *Biological Journal of the Linnean Society*, **103**, 779–800.

Kirkpatrick M, Hall DW (2004) Sexual selection and sex linkage. *Evolution*, **58**, 683–691.

Larson EL, White TA, Ross CL, Harrison RG (2014) Gene flow and the maintenance of species boundaries. *Molecular Ecology*, **23**, 1668–1678.

Lawniczak MKN, Emrich SJ, Holloway AK *et al.* (2010) Widespread divergence between incipient *Anopheles gambiae* species revealed by whole genome sequences. *Science*, **330**, 512–514.

Lexer C, Buerkle CA, Joseph JA, Heinze B, Fay MF (2007) Admixture in European *Populus* hybrid zones makes feasible the mapping of loci that contribute to reproductive isolation and trait differences. *Heredity*, **98**, 74.

Llopart A (2012) The rapid evolution of X-linked male-biased gene expression and the large-X effect in *Drosophila yakuba*, *D. santomea*, and their hybrids. *Molecular Biology and Evolution*, **29**, 3873–3886.

Loureiro A, Ferrand de Almeida N, Carretero MA, Paulo OS (2008) Atlas dos anfíbios e répteis de Portugal. *Instituto da Conservação da Natureza e da Biodiversidade, Lisboa*, **257**.

Luttikhuizen PC, Drent J, Peijnenburg K, Van der Veer HW, Johannesson K (2012) Genetic architecture in a marine hybrid zone: comparing outlier detection and genomic clines analysis in the bivalve *Macoma balthica*. *Molecular Ecology*, **21**, 3048–3061.

Lynch M, Force AG (2000) The origin of interspecific genomic incompatibility via gene duplication. *The American Naturalist*, **156**, 590–605.

Mallet J (2005) Hybridization as an invasion of the genome. *Trends in Ecology and Evolution*, **20**, 229–237.

Mallet J, Barton N, Lamas G *et al.* (1990) Estimates of selection and gene flow from measures of cline width and linkage disequilibrium in heliconius hybrid zones. *Genetics*, **124**, 921–936.

Matute DR, Butler IA, Turissini DA, Coyne JA (2010) A test of the snowball theory for the rate of evolution of hybrid incompatibilities. *Science*, **329**, 1518–1521.

Melo-Ferreira J, Alves PC, Freitas H, Ferrand N, Boursot P (2009) The genomic legacy from the extinct *Lepus timidus* to the three hare species of Iberia: Contrast between

mtDNA, sex chromosomes and autosomes. *Molecular Ecology*, **18**, 2643–2658.

Michel AP, Sim S, Powell THQ *et al.* (2010) Widespread genomic divergence during sympatric speciation. *Proceedings of the National Academy of Sciences*, **107**, 9724-9729.

Mullen SP, Dopman EB, Harrison RG (2008) Hybrid zone origins, species boundaries, and the evolution of wing-pattern diversity in a polytypic species complex of North American admiral butterflies (Nymphalidae: *Limenitis*). *Evolution*, **62**, 1400–1417.

Muller H (1942) Isolating mechanisms, evolution, and temperature. In: *Biology Symposium*, pp. 71–125.

Nakazato T, Jung M-K, Housworth EA, Rieseberg LH, Gastony GJ (2007) A genome-wide study of reproductive barriers between allopatric populations of a homosporous fern, *Ceratopteris richardii*. *Genetics*, **177**, 1141–1150.

Nosil P, Feder JL (2012) Genomic divergence during speciation: causes and consequences, **367**, 332–342.

Nosil P, Harmon LJ, Seehausen O (2009) Ecological explanations for (incomplete) speciation. *Trends in Ecology & Evolution*, **24**, 145–156.

Orr HA (1987) Genetics of male and female sterility in hybrids of *Drosophila pseudoobscura* and *D. persimilis*. *Genetics*, **116**, 555–563.

Parchman TL, Gompert Z, Braun MJ *et al.* (2013) The genomic consequences of adaptive divergence and reproductive isolation between species of manakins. *Molecular Ecology*, **22**, 3304–3317.

Parchman TL, Gompert Z, Mudge J *et al.* (2012) Genome-wide association genetics of an adaptive trait in lodgepole pine. *Molecular Ecology*, **21**, 2991–3005.

Payseur BA (2010) Using differential introgression in hybrid zones to identify genomic regions involved in speciation. *Molecular Ecology Resources*, **10**, 806–820.

Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE (2012) Double digest RADseq: an inexpensive method for *de novo* SNP discovery and genotyping in model and non-model species. *PloS One*, **7**, e37135.

Pinho C, Kaliontzopoulou A, Carretero MA, Harris DJ, Ferrand N (2009) Genetic admixture between the Iberian endemic lizards *Podarcis bocagei* and *Podarcis carbonelli*: Evidence for limited natural hybridization and a bimodal hybrid zone. *Journal of Zoological Systematics and Evolutionary Research*, **47**, 368–377.

Pritchard JK, Pickrell JK, Coop G (2010) The genetics of human adaptation: hard

sweeps, soft sweeps, and polygenic adaptation. *Current Biology*, **20**, R208–R215.

Purcell J, Brelsford A, Wurm Y, Perrin N, Chapuisat M (2014) Convergent genetic architecture underlies social organization in ants. *Current Biology*, **24**, 2728–2732.

Qvarnström A, Bailey RI (2009) Speciation through evolution of sex-linked genes. *Heredity*, **102**, 4.

Ravinet M, Faria R, Butlin RK *et al.* (2017) Interpreting the genomic landscape of speciation: a road map for finding barriers to gene flow. *Journal of Evolutionary Biology*, **30**, 1450–1477.

Ribeiro MJMC (2014) A landscape genetics perspective on the spatial dynamics of hybridization between two species of wall lizards. MSc thesis in Biodiversity, Genetics and Evolution, Faculty of Sciences, University of Porto.

Rieseberg LH (2009) Evolution: replacing genes and traits through hybridization. *Current Biology*, **19**, R119–R122.

Ritchie MG (2007) Sexual selection and speciation. *Annual Review of Ecology, Evolution, and Systematics*, **38**, 79–102.

Sæther SA, Sætre G-P, Borge T *et al.* (2007) Sex chromosome-linked species recognition and evolution of reproductive isolation in flycatchers. *Science*, **318**, 95–97.

Sætre G, Borge T, Lindroos K *et al.* (2003) Sex chromosome evolution and speciation in *Ficedula* flycatchers. *Proceedings of the Royal Society of London B: Biological Sciences*, **270**, 53–59.

Scascitelli M, Whitney KD, Randell RA *et al.* (2010) Genome scan of hybridizing sunflowers from Texas (*Helianthus annuus* and *H. debilis*) reveals asymmetric patterns of introgression and small islands of genomic differentiation. *Molecular Ecology*, **19**, 521–541.

Sillero N, Brito JC, Skidmore AK, Toxopeus AG (2009) Biogeographical patterns derived from remote sensing variables: the amphibians and reptiles of the Iberian Peninsula. *Amphibia-Reptilia*, **30**, 185–206.

Storchova R, Gregorová S, Buckiova D *et al.* (2004) Genetic analysis of X-linked hybrid sterility in the house mouse. *Mammalian Genome*, **15**, 515–524.

Storchová R, Reif J, Nachman MW (2010) Female heterogamety and speciation: Reduced introgression of the z chromosome between two species of nightingales. *Evolution*, **64**, 456–471.

Strasburg JL, Sherman NA, Wright KM *et al.* (2012) What can patterns of differentiation across plant genomes tell us about adaptation and speciation? *Philosophical Transactions of the Royal Society B: Biological Sciences*, **367**, 364–373.

Szymura JM, Barton NH (1986) Genetic analysis of a hybrid zone between the fire-bellied toads, *Bombina bombina* and *B. variegata*, near Cracow in southern Poland. *Evolution*, **40**, 1141–1159.

Szymura JM, Barton NH (1991) The genetic structure of the hybrid zone between the fire-bellied toads *Bombina bombina* and *B. variegata*: comparisons between transects and between loci. *Evolution*, **45**, 237–261.

Tarroso P, Pereira RJ, Martínez-Freiría F, Godinho R, Brito JC (2014) Hybridization at an ecotone: Ecological and genetic barriers between three Iberian vipers. *Molecular Ecology*, **23**, 1108–1123.

Teeter K, Payseur B (2008) Genome-wide patterns of gene flow across a house mouse hybrid zone. *Genome Research*, **18**, 67–76.

Teeter KC, Thibodeau LM, Gompert Z *et al.* (2010) The variable genomic architecture of isolation between hybridizing species of house mice. *Evolution: International Journal of Organic Evolution*, **64**, 472–485.

Templeton AR (1992) The meaning of species and speciation: a genetic perspective. *The Units of Evolution*, 159–185.

Tucker PK, Sage RD, Warner J, Wilson AC, Eicher EM (1992) Abrupt cline for sex chromosomes in a hybrid zone between two species of mice. *Evolution*, **46**, 1146–1163.

Turner TL, Hahn MW, Nuzhdin S V (2005) Genomic islands of speciation in *Anopheles gambiae*. *PLoS Biology*, **3**, e285.

Via S (2012) Divergence hitchhiking and the spread of genomic isolation during ecological speciation-with-gene-flow. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, **367**, 451–460.

Wu CI (2001) The genic view of the process of speciation. *Journal of Evolutionary Biology*, **14**, 851–865.

Yatabe Y, Kane NC, Scotti-Saintagne C, Rieseberg LH (2007) Rampant gene exchange across a strong reproductive barrier between the annual sunflowers, *Helianthus annuus* and *H. petiolaris*. *Genetics*. **175**,1883-1893.

Article III. Evolution of sympatry without complete reproductive isolation: does hybridization matter for *Podarcis carbonelli* conservation?

Guilherme Caeiro-Dias[1,2,3], Alan Brelsford[4,5], Pierre-André Crochet[3], Catarina Pinho[1]

[1]Centro de Investigação em Biodiversidade e Recursos Genéticos, CIBIO/InBio, Vairão, Portugal
[2]Departamento de Biologia, Faculdade de Ciências, Universidade do Porto, Porto, Portugal
[3]CEFE, CNRS, University of Montpellier, University Paul Valéry Montpellier 3, EPHE, IRD, Montpellier, France
[4]Department of Ecology and Evolution, University of Lausanne, Lausanne, Switzerland
[5]Biology Department, University of California Riverside, California, USA

Abstract

Studies on speciation have traditionally been focused on the development of reproductive isolation. However, a significant number of studies have demonstrated that speciation with gene flow is relatively common in nature. In conservation biology, interspecific gene flow raises problematic questions. Discontinuities between species evolve gradually, and defining particular species or conservation units in contexts where gene flow occurs is not always obvious. Also, hybridization raises concerns about its negative effects on the conservation of endangered species and is especially problematic for rare species that contact with other, more abundant, species. However, in the last years, introgressive hybridization has also been growingly recognized as a source of diversity and new advantageous alleles. Carbonell's wall lizard (*Podarcis carbonelli* Pérez-Mellado 1981) is an endangered species whose distribution overlaps everywhere at least with one of five other taxa. Here we want to determine whether *P. carbonelli* is completely reproductively isolated from other congeneric species and to evaluate whether hybridization is a phenomenon that needs to be taken into account when developing a conservation plan. We used restriction site associated DNA (RAD) sequencing to discover SNPs on samples from four contact zones between *P. carbonelli* and other four species. Principal component analysis, multilocus genotype assignment and measures of interspecific heterozygosity suggest

that reproductive isolation is still incomplete but major barriers to gene flow are evident in all analyzed hybrid zones. While species cohesion does not seem a major problem for *P. carbonelli*, the indirect effects of hybridization is, such as the waste of reproductive effort. Additionally the consequences of complex interactions with other threats, as current climatic changes or possible interchanges of beneficial alleles, emphasize the clear need to take into account such relationships for conservation strategies. Urgent conservation measures are needed for this species and here we anticipate the need to evaluate the consequences of hybridization in distinct populations and geographic contexts and include that information in a comprehensive conservation plan for *P. carbonelli*.

Key words: reproductive isolation; hybridization; patterns of introgression; conservation; *Lacertidae*; RAD sequencing.

Introduction

Speciation is the evolution of reproductive isolation (Mayr 1942, von Holdt *et al.* 2018). Nonetheless, recent advances in molecular genetics have led to an increasing number of reports about hybridization and introgression between fully recognized species, emphasizing limitations to this concept and demonstrating that complete reproductive isolation cannot be taken as a prerequisite for species delimitation. A lot of species are maintained in nature either coexisting with different niche, phenotypes, etc. (e.g. Whittemore & Schaal 1991; Milne *et al.* 1999; Neaves *et al.* 2010; Hochkirch & Lemke 2011) or replacing each other abruptly at narrow contact zones (e.g. Szymura & Barton 1986; Irwin *et al.* 2009; Tarroso *et al.* 2014; Grossen *et al.* 2016) and still hybridize and exchange genes. Moreover, human-mediated species translocations and range modifications have confirmed that allopatric species often hybridize when in contact (e.g. Rubidge & Taylor 2005; Ayres *et al.* 2008; Senn & Pemberton 2009). Thus, it has been increasingly accepted that speciation without complete reproductive isolation can happen even in complete allopatry with no gene flow during the speciation process. This perspective demands a new outlook on gene flow in fields like evolutionary biology, ecology or conservation biology.

The two key questions when diverging populations occur in sympatry or parapatry and interbreed are: i) how frequent is hybridization and ii) wheather hybridization also occurs in later backcross generations leading to introgression. In conservation biology, along with the obvious problem in delimiting conservation units, gene flow raises problematic questions. Distinct views of evolutionary consequences of hybridization often come into clash (Butlin & Ritchie 2013; Sætre 2013), which is translated into difficulties to make management decisions (Allendorf *et al.* 2001; Wayne & Shaffer 2016; vonHoldt *et al.* 2018). On one hand, hybridization is often regarded as having negative effects on the conservation of endangered species. Indeed, extensive introgression may decrease divergence (Seehausen *et al.* 2008) and ultimately lead to genetic swamping (e.g. Rhymer *et al.* 1994). There are also indirect negative effects of hybridization related to the waste of reproductive effort in the generation of unviable or maladapted hybrid offspring (Lepais *et al.* 2009; Beatty *et al.* 2010) or to the loss of locally adapted alleles (Bourret *et al.* 2011). In extreme cases these situations may contribute to extinction (McMillan and Wilcove 1994), particularly in species that are already rare or affected by other kinds of threats. On the other hand, in recent years hybridization and introgression have been growingly recognized as a source of genetic novelty, increasing diversity and prompting the flow of advantageous alleles (Anderson *et al.* 2009; Whitney *et al.* 2010; Becker *et al.* 2013), promoting for example an increase of resilience of endangered populations (Tompkins *et al.* 2006) and even triggering biological diversification, all of which suggest that the effects of hybridization on endangered species may be overall positive and encouraged.

Carbonell's wall lizard (*Podarcis carbonelli* Pérez-Mellado 1981) is a species currently listed as endangered by IUCN (Sá-Sousa *et al.* 2009) due to its small extent of occurrence, fragmentation and declining population trend. This species is a small lacertid lizard belonging to the *Podarcis hispanicus* species complex, a clade of closely-related species inhabiting the Iberian Peninsula and North Africa (Kaliontzopoulou *et al.* 2011b). *Podarcis carbonelli* is endemic to the western Iberian Peninsula where is has a very fragmented distribution in western Portugal, west-central Spain and very small areas of south-western Spain (Sá-Sousa 2008). Moderate levels of population substructure (in the context of what is typical in *Podarcis*) were found in genetic variation assessments, probably resulting from some level of isolation between populations (Pinho *et al.* 2007, 2011). The present geographic distribution of *P. carbonelli* has been

shaped by an important range reduction due to climatic changes after the last glacial maximum (Sá-Sousa 2001; Sillero & Carretero 2013) and a current declining trend has been detected by field studies in recent years (Sillero *et al.* 2012, 2014). The species is present in several protected areas throughout Iberia, which is expected to guarantee some natural habitat preservation, but there is no official conservation plan directly aimed at this species.

Podarcis carbonelli has an exceptional distribution pattern in the context of the *P. hispanicus* complex. While most species of the *P. hispanicus* complex replace each other abruptly or meet in very narrow contact zones, a pattern that is likely mediated by competitive exclusion (Caeiro-Dias *et al.* 2018), *P. carbonelli* overlaps with at least five other taxa: *P. bocagei, P. vaucheri*, *P. guadarramae guadarramae, P. g. lusitanicus* and *P. virescens.* With the first two species, the overlapping areas are very restricted (Carretero *et al.* 2002; Sá-Sousa & Harris 2002; Pinho *et al.* 2009), but with the other three, the areas of sympatry are more extensive (Caeiro-Dias et al. 2018; see Figure 3.2.1.). Natural hybridization and introgression have been widely documented in *Podarcis* (Capula 1993, 2002), and the *P. hispanicus* complex makes no exception (Pinho *et al.* 2009; Renoult *et al.* 2009). *Podarcis carbonelli* itself is known to hybridize with *P. bocagei* in a narrow contact zone in northwestern Portugal (Pinho *et al.* 2009, Caeiro-Dias *et al.* in prep.) but nothing is known about the level of reproductive isolation with the other species it coexists with. This widespread contact and the worrisome conservation status of *P. carbonelli* thus raise the question of how prevalent is the exchange of genes between *P. carbonelli* and its congeners. On one hand, the evidence for genetic permeability of *Podarcis* species raise the possibility that gene flow is indeed occurring frequently, as in most species pairs of the *Podarcis hispanicus* complex studied to date. On the other hand, it is expected that hybridization do not last long when species establish large-scale sympatry, for instance processes like reinforcement may enhance the effects of barriers to gene flow in less divergent taxa when in sympatry (Coyne & Orr 1997; Boughman 2001; Yukilevich 2012), decreasing the levels of admixture. The entirely sympatric distribution of *P. carbonelli* suggests that mate-recognition mechanisms should be more efficient than in the parapatric contacts studied so far.

In this study we have two main objectives. Firstly we want to determine the extent of *P. carbonelli* hybridization when in sympatry with other species and if it leads to

interspecific gene flow. Secondly we want to use data on gene flow to evaluate whether hybridization with congeners is a phenomenon that needs to be taken into account when devising a conservation plan for *P. carbonelli*. To do so, we collected genetic data from areas of syntopy and analyzed patterns of hybridization and introgression between *P. carbonelli* and four of the five co-distributed lineages using single nucleotide polymorphisms (SNPs) detected with restriction site associated DNA (RAD) sequencing (Baird *et al.* 2008; Hohenlohe *et al.* 2010). In the presence of recurrent backcrosses we can demonstrate the lack of complete reproductive isolation. This aspect should be taken into account along with other sources of information for developing an appropriate conservation strategy for this endangered species.

Material and Methods

Sampling

Samples were collected between spring and autumn of 2013 in four contact zones, all very restricted in space (Figure 3.2.1.). In all contact zones, the two species were found in actual syntopy. The syntopy area between *P. virescens* and *P. carbonelli* was the castle of Santa Maria da Feira (40.921 N / -8.543 W, Aveiro District, Portugal) where the final dataset contained 26 individuals identified in the field as *P. virescens* and 23 as *P. carbonelli*. The area of syntopy with *P. guadarramae lusitanicus* was a 500 meters long area in Vale do Rossim (40.403 N / -7.587 W, Serra da Estrela Natural Park, Guarda District, Portugal), an area dominated by sparse pine trees with relatively dense scrub cover and rocky outcrops where 38 samples were identified as *P. guadarramae lusitanicus* and 17 as *P. carbonelli* in the field. Specimens in syntopy with *P. vaucheri* were found in a few dozens of meters in the extreme northwest of the coastal village of Matalascañas (37.005 N / -6.566 W, Huelva Province, Spain) where 50 specimens were identified as *P. vaucheri* and 10 as *P. carbonelli*. Data from the contact zone with *P. bocagei* comes from a narrow coastal dune stripe with dune scrub vegetation north of Espinho (41.020 N / -8.643 W, Aveiro District, Portugal, see Pinho *et al.* 2009, Caeiro-Dias *et al.* in prep). Fifteen samples identified as *P. bocagei* and 100 identified as *P. carbonelli* were found in an area about 450 meters wide where actual syntopy was observed. In Matalascañas, both species inhabited distinct micro-habitat (*P. vaucheri* on human-made structures, *P. carbonelli* in semi-natural dune environments with pine trees

and scrub vegetation), but in other syntopy area no ecological segregation was apparent.

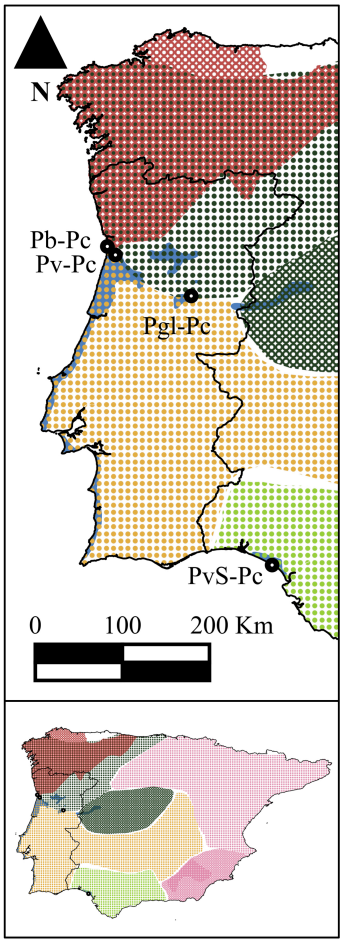

Figure 3.2.1. Distribution of each *P. hispanicus* complex species in the Iberian Peninsula (bottom map) with the regions where the contact zones were sampled in evidence (top map). The species included in this study are *Podarcis carbonelli* (Pc, shaded blue areas), *P. bocagei* (Pb, red circles), *P. virescens* (Pv, yellow points), *P. guadarramae lusitanicus* (Pgl, dark green points), and *P. vaucheri* (PvS, light green points). *P. guadarrame lusitanicus*, with which *P. carbonelli* overlaps but no contact zone was sampled for this study, is also represented (dark green circles). Each black point represents the location of the contact zones studied here and is identified with the acronyms of the contacting species.

In all contact zones the sampling scheme aimed to catch all the individuals that were seen, avoiding bias towards species, sex or age. All the information about the samples is summarized in Appendix III, Table S3.1. Lizards were captured by noose around the neck, which is harmless, and kept in individual cloth bags until they were processed. All the samples were identified to species in the field based mostly on a combination of habitus, coloration and head shape, geo-referenced and photographed. None of the individuals were identified as intermediate phenotypes between parental species. A small tail tip was collected and immediately stored in 96% ethanol for subsequent DNA extraction. Animals were liberated in the same day in the place of capture. We added reference individual samples from outside the contact zones retrieved from the tissue collections of the CIBIO-InBio, Portugal and EPHE-CEFE,

France. We used 20 *P. bocagei* samples as reference, 12 *P. virescens*, 37 *P. g. lusitanicus*, 15 *P. vaucheri* and 41 *P. carbonelli* (see Appendix III, Table S3.1. for details). These samples were distributed across the range of each species to capture as much of its diversity as possible. In the case of the *P. bocagei* and *P. carbonelli* contact zone, we included as reference samples from the ends of the transect analyzed by Caeiro-Dias *et al* (in prep.).

RAD sequencing, data filtering and SNP calling

We obtained RAD sequencing data using modifications to protocols from Parchman *et al.* (2012), Peterson *et al.* (2012) and Purcell *et al.* (2014). The complete protocol is described by Brelsford *et al.* (2016). The main steps were the digestion of genomic DNA with the restriction enzymes *Sbf*I and *Mse*I, ligation of barcoded adapters to restriction sites, amplification of each individual sample in four independent separate PCR reactions, pool of all PCR products and fragment selection between 400 and 500 bp using a 2.5% agarose gel. Samples used in this work were sequenced from two distinct libraries; one had a total of 665 samples and the other 749 (including other samples not used in this study). One library was sequenced on two Illumina® (San Diego, CA, USA) HiSeq 2000 lanes at the Lausanne Genetic Technology Facility (Lausanne, Switzerland) and on four Illumina® (San Diego, CA, USA) HiSeq 1500 lanes at the CIBIO Next Generation Sequencing Platform (Vairão, Portugal), with single-end 100 bp reads. The other was sequenced on four lanes at CIBIO. The dataset for the contact zone between *P. bocagei* and *P carbonelli* was obtained from a previously prepared library (see Caeiro-Dias *et al* in prep.).

We demultiplexed individual raw reads using the process_radtags module of Stacks version 2.0 beta8 (Catchen 2013) allowing one mismatch per barcode, to remove low-quality reads, reads containing adapter sequence, reads with uncalled bases and reads that failed the Illumina® 'chastity' filter. We then tested the optimal *de novo* assembly parameters for our data set following the protocol described in Rochette & Catchen (2017) adapted to Stacks version 2.0, prior to final *de novo* read alignment. For this test we used the samples of one contact zone only (*P. carbonelli* x *P. vaucheri*), running consecutively ustacks (build loci), cstacks (create a catalogue of loci), sstacks (match individual samples against the catalogue), tsv2bam (transpose data), gstacks (align each read to a locus and call SNPs) and populations (SNP filtering and output

data) units. We performed distinct runs varying the number of mismatches allowed between reads within individuals (*M*) and between individuals (*n*). As suggested by Rochette and Catchen (2017), we varied *M* and *n* between 1 and 9 and keeping *M* = *n*, while the minimum depth of coverage to accept a stack (*m*) was kept constant as 3, the ustacks default. Both in these prior tests and in the final analyses we used a bounded SNP model (--model_type bounded option) in ustacks, with an upper bound for the error rate of 0.1. Besides the mentioned changes and tests to parameter values, we used default settings for all other steps of the pipeline. After completion of the pipeline for the 9 different sets of parameter values, we analysed the numbers of loci shared by at least 80% of the samples and the distribution of variable sites within loci for the range of tested values. Based on this analysis we chose to retain *M* = *n* = 5 for the final *de novo* read alignment and SNP calling procedure, which was conducted in a separate way for each contact zone but including the references for each one.

For the three datasets sequenced specifically for this study, the populations module from Stacks was used to filter out the resulting variants with more than 0.7 maximum observed heterozygosity and to keep only one SNP per locus chosen at random. Subsequently, vcftools version 0.1.15 (Danecek *et al.* 2011) was used to discard loci with depth coverage less than 8 and present in less than 35% of the samples, similarly to what was previously done for the dataset including samples from the contact zone between *P. bocagei* and *P. carbonelli*.

All the processes (library construction, demultiplexing and data filtering) were repeated independently for ~ 20% of the samples to evaluate replicability.

Genetic characterization of the contact zones and admixture analysis

Genomic variability among individuals was visualised by performing principal component analysis (PCA) on the dataset for each contact zone separately using the adegenet R package version 2.0.1 (Jombart 2008; Jombart & Ahmed 2011). We used Structure version 2.3.4 (Pritchard *et al.* 2000) to evaluate the proportion (*Q*) of each individual's genome originating from each of the parental species. We ran Structure for K=2 since we were only interested in detecting admixture between the two species. For clarity we will refer to the proportion of *P. carbonelli* (Q_C) in each contact zone, as it is the common element to all. The proportion of assignment of each individual to the other

species in a particular contact zone is 1 – Q_C. Structure was executed with 500 000 repetitions and a burn-in of 200 000. For Structure we included the reference individuals.

Individuals from reference populations outside the contact zones were then used as references to calculate the hybrid index (HI) of each individual from the admixed population using the R package Introgress version 1.2.3 (Gompert & Buerkle 2010). *P. carbonelli* parental individuals were set to have a HI of 0, and the other species were set to an HI of 1 in each contact zone. The proportion of loci in an admixed individual's genome with alleles inherited from both parental species, i.e. interspecific heterozygosity (*Het*), was calculated for each admixed individual using Introgress. This method for calculating interspecific heterozygosity assumes that parental allele frequencies are known. Therefore, the same individuals used as parentals for HI estimation were also used to calculate *Het*. A triangle plot can represent the relationship between the HI and *Het* (Fitzpatrick 2012). *Het* ranges from 0 to 1, whereas values near to 1 are interpreted as F1 hybrids and values lower than 1 indicate later generation hybrids that have either backcrossed with the parentals (overlapping the lines of the triangle plot) and/or with other hybrids (below the triangle plot lines; Fitzpatrick 2012). This analysis requires that loci are fully diagnostic between species or at least loci with large differences for a good approximation in the estimates. For three of the contact zones we performed this analysis by restricting the SNPs just to loci with fixed differences between reference individuals. However, in the case of the contact zone with *P. guadarramae lusitanicus*, where we sequenced fewer loci, we detected only one diagnostic locus between references. For this reason we performed also another analysis with loci with allelic frequencies higher than 0.75 in one species and lower than 0.25 in the other (“75/25” dataset). For the data to be comparable, we extended the analysis with a “75/25” dataset for the remaining contact zones.

To test whether the hybridization rate was effectively distinct between contact zones, i.e. recent generation hybrid composition, or alternatively if the proportions of recent generation admixed genotypes are similar across contact zones, we tested the independence of hybrid genotypic composition between contact zones by performing a pairwise Fisher’s exact test with R package stats version 3.3.3 (R Core Team, 2017) and then applying a Bonferroni correction for multiple comparisons. For each contact zone we classified individuals in two categories: parental individuals and later generation

backcrosses (Q_C < 0.2 and Q_C > 0.8) and recent admixed genotypes (Q_C between 0.2 and 0.8).

Results

Data filtering and SNP calling

The final datasets, after removing loci with depth coverage <8 and missing data >35%, and removing individuals with more than 35% of missing data, consisted of 8096 SNPs for *P. virescens* x *P. carbonelli*, 1085 for *P. guadarramae lusitanicus* x *P. carbonelli* and 5917 for *P. vaucheri* SSp x *P. carbonelli*. For *P. bocagei* x *P. carbonelli*, the final dataset consisted of 8201 SNPs, after removing loci with depth coverage <8, loci with missing data >20%, and individuals with missing data >30% (Caeiro-Dias *et al* in prep). The average coverage across individuals was 28.7, 24.8, 22.4 and 27.8 and across loci was 28.7, 27.3, 24.7 and 33.0 for *P. bocagei* x *P. carbonelli*, P*. virescens* x *P. carbonelli*, *P. guadarramae* x *P. carbonelli* and *P. vaucheri* SSp x *P. carbonelli*, respectively. Additionally, the "75/25" and the diagnostic data sets contained for each contact zones, respectively, 1362 and 714 SNPs in *P. bocagei* x *P. carbonelli*, 137 and 41 in *P. virescens* x *P. carbonelli*, 15 and 1 in *P. guadarramae lusitanicus* x P. carbonelli, 74 and 18 in *P. vaucheri* SSp x *P. carbonelli*.

Podarcis bocagei x *Podarcis carbonelli*

Despite the slightly different data set, our results replicate those obtained by Caeiro-Dias *et al.* (in prep), which was the source of the data used. The PCA analysis based on the 8201 SNPs focused on the syntopy area (15 *P. bocagei* and 100 *P. carbonelli*) separated two groups: PC1 explains 35.7% of the variance and separates *P. bocagei* (with less samples) from *P. carbonelli* (with more samples; Figure 3.2.2.a) but with several individuals in-between. The analysis with Structure was also consistent with those results showing several individuals with distinct proportions of the genome assigned to both species, revealing the occurrence of hybridization and introgression (Figure 3.2.3.a). Within the contact zone, Q_C was lower than 0.1 in 13 individuals and higher than 0.9 in 90 individuals. Q_C varied between 0.2 and 0.8 in 10 individuals and between 0.4 and 0.6 in six.

Estimates of *Het* were concordant between the "25/75" and diagnostic datasets (Figures 3.2.4.a and S3.1.a, respectively). Individuals with HI close to 0 and 1 had low estimates of *Het*. Individuals with intermediate HI had increased estimates of *Het* compared to reference individuals, i.e. were of mixed ancestry. Some individuals around to triangle plot lines (or in the triangle plot line for the diagnostic data set) suggest backcrosses between hybrids and one of the parental species, while individuals above the line, i.e. with intermediate and low *Het* and HI deviated from 0, 1 and 0.5, indicate crosses between individuals with mixed ancestry (Figures 3.2.4.a and S3.1.a.).

Podarcis virescens x *Podarcis carbonelli*

The PCA focused in the contact zone (26 *P. virescens* and 23 *P. carbonelli*) showed two clusters of individuals based on the 8096 SNPs analysed (Figure 3.2.2.b). PC1 explains 39.2% of the variation and separates the two species, PC2 explains 3.1% of the variation and captures intraspecific variability in *P. virescens*. The assignment analysis is concordant with the results of the PCA (Figure 3.2.3.b). Q_C was 1 or 0 for all but two individuals (Q_C = 0.03 and 0.04) for K = 2. Both analyses of *Het* were also concordant with previous results. Therefore no hybridization or introgression was detected in this contact zone.

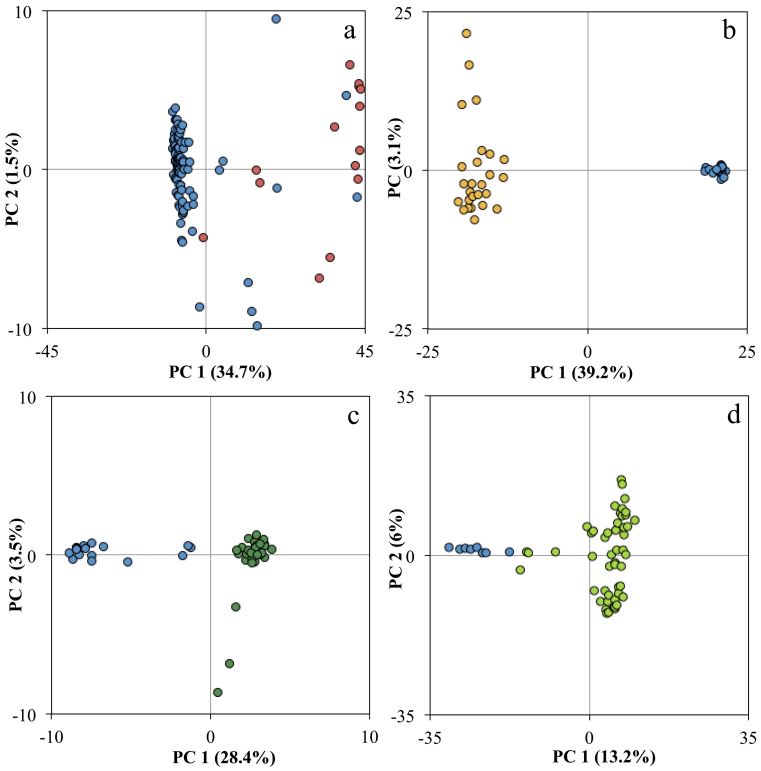

Figure 3.2.2. Principal Component Analysis of each SNP dataset variation in the contact zones between **a)** *P. bocagei* x *P. carbonelli*, **b)** *P. virescens* x *P. carbonelli*, **c)** *P. guadarramae lusitanicus* x *P. carbonelli* and **d)** *P. vaucheri* SSp x *P. carbonelli*. Individual colours (as in Figure 3.2.1.) based on morphological identification. The variation explained by each axis (PC) is represented as percentage.

Podarcis guadarramae lusitanicus x *Podarcis carbonelli*

The PCA on the samples from the contact zone (38 *P. guadarramae lusitanicus* and 17 *P. carbonelli*) revealed two main groups of samples based on the 1085 SNPs (28.4% of the variance explained by PC1) showing that the two species in the contact zone are clearly separable but few individuals appear in an intermediate position between both groups (Figure 3.2.2.c). Furthermore, PC2 explains 3.5% of the variation and captures intraspecific variability within *P. g. lusitanicus*. Structure returned 36 individuals with Q_C lower than 0.1, 15 with Q_C higher 0.9 and three had Q_C close to 0.5, suggesting some degree of mixed ancestry (Figure 3.2.3.c).

Estimates of *Het* were low for most individuals (Figure 3.2.4.c). Despite the large dispersion of *Het* value for individuals with HI close to 0 and 1, likely due to the lack of diagnostic markers, two individuals with Q_C close to 0.55 in the Structure analysis had HI around 0.5 and *Het* close to 1, which is typical of F1 hybrids. Furthermore, one individual with $Q_C \approx 0.6$ had HI ≈ 0.4 and *Het* ≈ 0.8 suggesting a backcross between an individual with mixed ancestry and a *P. carbonelli* parental. However, given the Structure results, it can rather be an artefact due to the lack of fully diagnostic loci in the estimates of the *He* and HI.

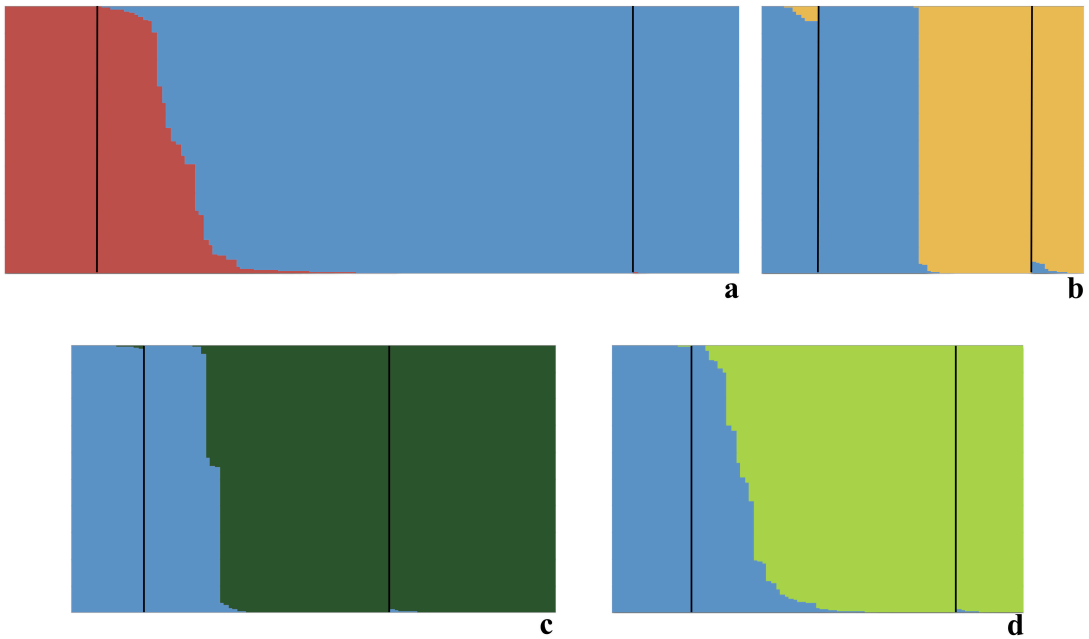

Figure 3.2.3. Results from individual multilocus genotype clustering analysis for each contact zone between **a.** *P. bocagei* x *P. carbonelli*, **b.** *P. virescens* x *P. carbonelli*, **c.** *P. guadarramae lusitanicus* x *P. carbonelli* and **d.** *P. vaucheri* SSp x *P. carbonelli*. Each individual is represented as a vertical line partitioned into the K =2 colored segments, whose length is proportional to the K colours. Black lines delimit individuals from contact zone (between the two vertical lines) from the reference samples (left and right of the vertical lines). Colors as in Figure 3.2.1.

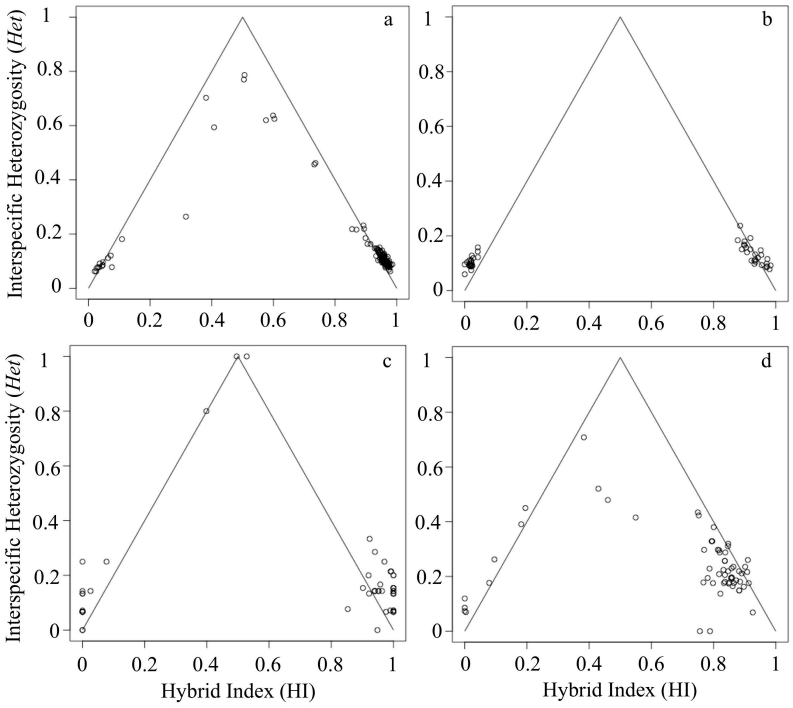

Figure 3.2.4. Distributions of individual hybrid index and interspecific heterozygosity in the contact zones between **a.** *P. bocagei* x *P. carbonelli*, **b.** *P. virescens* x *P. carbonelli*, **c.** *P. guadarramae lusitanicus* x *P. carbonelli* and **d.** *P. vaucheri* SSp x *P. carbonelli* based on loci with allelic frequencies higher than 0.75 in one species and 0.25 in the other. In each contact zone *P. carbonelli* reference individuals were set to have a HI of 0, and the other species was set to an HI of 1.

Podarcis vaucheri x *Podarcis carbonelli*

The first PCA axis (PC1) using the 5917 SNPs on the contact zone (50 *P. vaucheri* and 9 *P. carbonelli*) explained 13.2% of the variance (Fig. 3.2.3.d) and separated *P. carbonelli* from *P. vaucheri* with several individuals between the two groups. A large amount of variation among *P. vaucheri* individuals is apparent along PC2 (6% of the variance). Structure identified 11 individuals with Q_C between 0.1 and 0.9 and six between 0.2 and 0.8, four of which with Q_C between 0.4 and 0.6 (Figure 3.2.3.d).

In this contact zone we found fewer individuals with *HI* = 0 or 1 and *Het* ≈ 0 than would be expected if hybridization was restricted (Figure 3.2.4.d). However, when including only diagnostic loci, estimates of *Het* are 0 or close to zero for several individuals (Figure S3.1.c). Most individuals had *HI* between 0.75 and 0.93 with estimates of *Het* between 0 and 0.4 with the “25/75” dataset, suggesting that at least some individuals, particularly those with HI closer to 0.75, are the result of recurrent backcrosses between hybrids and *P. vaucheri* SSp. These results are not very different from the results with the diagnostic dataset. The two main differences are that with the diagnostic dataset we clearly identify the individuals with *Het* close to zero for both extreme HI values (0 and 1) and on the right side of the triangle plot we can distinguish the several backcrosses of admixed individuals with parental *P. vaucheri* Sp from the

backcrosses between hybrids. Both datasets show five individuals with *HI* and *Het* values that place them close to the left line of the triangle (Figure 3.2.4.d), indicating backcrosses of individuals from distinct hybrid classes with *P. carbonelli*.

Comparison between contact zones

Three out of the six comparisons between pairs of contact zones were not significant (*p*-value > $8.33x10^{-3}$ after Bonferroni correction) and thus we could not reject the null hypothesis that the proportions of admixed genotypes in each contact zone were similar (Table 3.2.1.). However, the other three comparisons were significant and we could reject the null hypothesis (*p*-value < $8.33x10^{-3}$). All the three significant comparisons involved the contact zone between *P. carbonelli* and *P. vaucheri* Sp. indicating that admixture in this contact zone was significantly higher than in the others (Figure 3.2.5.).

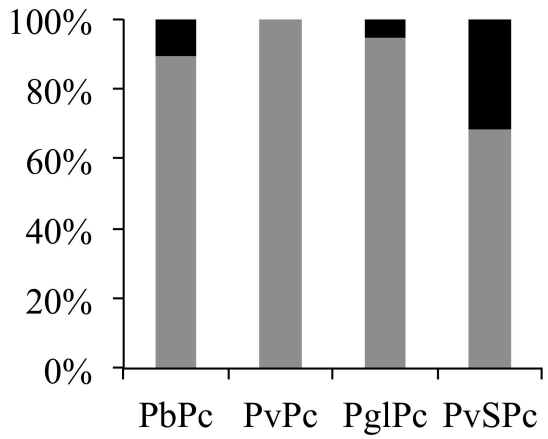

Figure 3.2.5. Proportion of parental and later generation hybrid genotypes resulting from recurrent backcrosses ($Q_C \geq 0.8$; grey) and earlier generation hybrid genotypes, ($Q_C < 0.8$; black) in each contact zone. PbPc: *P. bocagei* x *P. carbonelli*, PvPc *P. virescens* x *P. carbonelli*, PglPc *P. guadarramae lusitanicus* x *P. carbonelli*, PvSPc: *P. vaucheri* SSp x *P. carbonelli*.

Table 3.2.1. Fisher's exact test results on each of the six possible pairwise comparisons between contact zones studied. * indicates the significant tests after Bonferroni correction (significant at *p*-value < $8.33x10^{-3}$). Acronyms as in Figure 3.2.5.

Comparison	*p*-value
PbPc x PglPc	0.39
PbPc x PvPc	0.02
PbPc x PvSPc	$7.7x10^{-4}$*
PglPc x PvPc	0.24
PglPc x PvSPc	$3.15x10^{-4}$*
PvPc x PvSPc	$4.5x10^{-6}$*

Discussion

Our results provide evidence that *P. carbonelli*, a species co-distributed across most of its distribution range with at least another *Podarcis* species can persist in extensive sympatry even in the face of gene flow. In three of the four contact zones investigated we found that hybridization occurs and is not restricted to F1 hybrids. Reproductive isolation seems to be complete only between *P. carbonelli* and *P. virescens*. Thus, it is clear that hybridization cannot be neglected when developing a conservation plan for *P. carbonelli*.

Persistence of Podarcis *species in the face of gene flow*

The five species studied here are indubitably distinct, with divergence times based on mtDNA ranging between 4 Mya (*P. virescens* and *P. carbonelli*) and 10 Mya *(P. vaucheri* SSp and *P. carbonelli*; Kaliontzopoulou *et al.* 2011b), most of them with distinct morphology (Kaliontzopoulou *et al.* 2011a) and climatic niches (Caeiro-Dias *et al.* 2018). Still, our results suggest that hybridization is relatively common in three out of the four contact zones. The occurrence of hybridization events does not necessarily imply incomplete reproductive isolation. If F1 hybrids are unfit or are sterile, hybrid zones possess only parental and F1 genotypes (Allendorf *et al.* 2001; Steeves *et al.* 2010) and thus parental species have complete reproductive isolation. This may be a possible scenario between *P. carbonelli* and *P. guadarramae lusitanicus* since we found, most likely, only F1 hybrids. However we found evidence in two contact zones (*P. bocagei* x *P. carbonelli* and *P. vaucheri* SSp x *P. carbonelli*) for the presence of later generation hybrids, disclosing interspecific gene flow and a lack of complete reproductive isolation. In both scenarios of hybridization with or without introgression, *P. carbonelli* is sympatric with several species across an important range and hybridization is likely to occur across numerous populations. Remarkably, we did not detect hybridization between *P. carbonelli* and its sister species *P. virescens* (Kaliontzopoulou *et al.* 2011b). Given the results for other contact zones and the closer evolutionary relationship between these two species, one could expect to detect higher hybridization levels between this pair but our SNP data set clearly discriminates both species without identifying hybrids. The use of high density RADseq markers have been proven to effectively detect hybrids and introgression (Hohenlohe *et al.* 2011; Pujolar *et al.* 2014), particularly within the groups

of *Podarcis* lizards studied here even without restricting analysis to loci with fixed differences (Caeiro-Dias *et al.* in prep., this study), thus it is clear that the lack of hybrids is not a by-product of lack of detectability.

Although *P. carbonelli* is not completely reproductively isolated from other *Podarcis* species, our data suggest that there are mechanisms avoiding extensive introgression. Given that they all occurr in strict syntopy, that the contact zones are build up mostly by parental individuals and that they breed during the same period of the year, like other *Podarcis* species (Pérez-Mellado 1982; Carretero *et al.* 2006), there are no evident ecological or temporal barriers to gene flow in most contact zones and thus pre-zygotic barriers preventing gene flow are possibly an important mechanism of isolation across all contact zones. The most obvious example is that there is no spatial segregation at a micro scale that might be preventing hybridization between *P. carbonelli* and *P. virescens,* since we have previously reported one *P. virescens* adult male eating a *P. carbonelli* tail from a new-born individual in the exact same location (Dias *et al.* 2016; Appendix IV), both individuals included in this study. Therefore, these two species clearly meet and interact in this location, but do not hybridize. On the other hand, where we found more hybrids (*P. vauheri* SSp vs *P. carbonelli*) is where both species seem to have a higher spatial segregation.

Our results suggest the presence of barriers to extensive interspecific gene flow across several contact zones, otherwise it would be expected that hybridization led to extensive introgression (Seehausen *et al.* 2008). Given the overall deep divergence between these species, these observations are consistent with late stages of speciation, in which reproductive isolation is still incomplete but where major barriers to gene flow are strikingly evident and successful reproduction is halted between diverging populations. This is true for all the species pairs examined. Consequently, recurrent introgression and recombination do not seem to affect species cohesion.

Why considering hybridization in Podarcis carbonelli *for conservation plans?*

We detected hybridization between *P. carbonelli* and most of its co-occurring species. In other systems where endangered species hybridize this was recognized as a concern (Milián-García *et al.* 2015; Vuillaume *et al.* 2015), but given that overall *Podarcis* species studied here remain distinct in sympatry, globally genetic swamping does not seem to be a major current threat to *P. carbonelli*. A similar conclusion was achieved for

the critically endangered bird *Himantopus novaezelandiae* (Steeves *et al.* 2010). In the regions where *P. carbonelli* contacts with *P. bocagei* and *P. guadarramae lusitanicus*, hybridization and introgression seem to be relatively rare. However, while with *P. bocagei* introgression is very restricted in space (Caeiro-Dias *et al.* in prep.), interspecific gene flow may potentially occur across all the areas where *P. carbonelli* is sympatric with *P. guadarramae lusitanicus,* corresponding to a large area of the distribution. Similarly, interspecific gene flow may potentially occur across all the sympatric area between *P. carbonelli* and *P. vaucheri* SSp. The contact zone between these two species is particular in the sense that we found a significantly higher number of admixed genotypes.

The finding reported here raises the question of whether low *P. carbonelli*'s densities increase the rate of hybridization with other species. We show that levels of hybridization and introgression change depending on the contact zone. One possible explanation for such differences is the species relative abundance. The increased frequency of hybridization in one contact zone may be related with the high frequency of *P. vaucheri* SSp compared with *P. carbonelli*. This hypothesis remains to be tested and the detectability of both species may have influenced the highly skewed sampling towards *P. vaucheri* SSp. (we collected 50 *P. vaucheri* SSp against eight *P. carbonelli*). However, ecological niche modelling suggests that the isolated populations of *P. carbonelli* in southwest Spain do not inhabit in regions with optimal conditions for the species (Sillero & Carretero 2013; Caeiro-Dias *et al.* 2018) which may result in low abundances (Brown 1984). In this context, low abundance of *P. carbonelli* relatively to *P. vaucheri* SSp might increase the opportunities for interspecific mates of the first with the latter, increasing hybridization rate. This has been observed for other species where one of them is less abundant than the other (Burgess *et al.* 2005; Lepais *et al.* 2009; Beatty *et al.* 2010). Density-mediated hybridization gains even more importance since frequently *P. carbonelli* occurs in small fragmented populations, for example, in highlands (Sá-Sousa 2008). Indeed, the contact zones where we sampled less *P. carbonelli* individuals (the contact zones with *P. guadarramae lusitanicus* and *P. vaucheri* SSp) correspond to smaller and isolated populations. Moreover, the declining tendency has been recently observed in several populations (Sillero *et al.* 2012, 2014). Also, while *P. vaucheri* SSp was observed essentially in human made structures at Matalacañas, *P. carbonelli* was detected outside of the town in semi-natural

environments and the syntopy between these two species seems to be restricted only to the edge of the town. Human-mediated habitat modifications have been widely identified as factors contributing to hybridization in other systems (Levin *et al.* 1996; Allendorf *et al.* 2001). An additional threat to *P. carbonelli* is the predicted overall range contraction due to climate changes (Sillero & Carretero 2013). The effects of climate change are occurring at a time when many *P. carbonelli* populations are already under pressure and the effects of hybridization in this species may be higher – and more severe – in the future. Hoffman & Sgrò (2011) emphasized the increment of the current pressure's effects due to climate changes and that such pressure can affect the evolutionary processes by changing the way genes move around landscapes and by introducing novel genotypes into populations through hybridization. The possible indirect effects of hybridization emphasize the clear need to take such relationships into account for conservation strategies.

P. carbonelli is currently considered endangered by IUCN (Sá-Sousa et al. 2009). This is one of the most threatened vertebrate species in the Iberian Peninsula but, unlike other endemic vertebrate species with the same current conservation status (e.g. *Anaecypris hispanica*, *Lynx pardinus*) or even lower (e.g. *Alytes cisternasii, Aquila adalberti*), to date there are no conservation plans with an integrated assessment of extinction risk. According to the IUCN red list (Sá-Sousa *et al.* 2009) "many of the southern populations are protected (including in the Coto Doñana National Park). In central Portugal and Spain, some populations are in natural parks." On the other hand "the southern populations are almost certainly at risk from climate change" (Sá-Sousa *et al.* 2009). For a species listed as endangered, protecting only the habitat is clearly an insufficient conservation effort. Noteworthy, two of our sampling locations where we detected hybridization are located on protected areas. Populations inhabiting protected areas may benefit from some habitat protected but this does not guarantee the long-term persistence of the species or, as our results demonstrate, the prevention of interspecific introgression. Hybridization has been long emphasized as a threat to several species (Rhymer & Simberloff 1996; Allendorf *et al.* 2001; Wolf *et al.* 2001; Wayne & Shaffer 2016), as well as a source of favourable alleles or allelic combinations (Anderson *et al.* 2009; Whitney *et al.* 2010; Becker *et al.* 2013). In the particular case of *P. carbonelli,* issues like the consequences of interspecific matings and gene flow may be extremely important both when negative effects are an additional threat to populations (e.g. wasted

reproductive effort in interspecific mates with other congeneric species can threaten small and fragmented populations of *P. carbonelli*), or when positive effects may enhance fitness (e.g. introgression of new alleles in the genomic background of *P. carbonelli* positively selected in the new environmental conditions, like in climate change or human-altered scenarios). Levels of hybridization and the extent of gene flow may depend on many factors such as level of divergence (Funk *et al.* 2006; Pereira *et al.* 2011), geographic context of divergence, and demography (Lepais *et al.* 2009; Beatty *et al.* 2010). We show that *P. carbonelli* does not hybridize with the closest co-occurring species in the sampled contact zone but hybridizes more often with the more divergent. We also confirmed that *P. carbonelli* has the highest hybridization rate in the contact zone where it is more rare. Interestingly, when *P. carbonelli* occurs in similar proportions than the other species, we did not detect any hybrid. These scenarios highlight the need to extend the analyses presented here to other populations and geographic contexts in order to understand the factors driving the magnitude of hybridization in this system. This data should also be used when designing a conservation plan for this species.

Concluding remarks

In conclusion, our results show the presence of recurrent gene flow between *P. carbonelli* and three congeneric species, despite deep divergence between them, which may be a common phenomenon throughout the species distribution. Therefore hybridization cannot be dissociated from conservation strategies. This species is listed as endangered and urgent conservation measures are needed. Here we anticipate the need to assess the consequences of hybridization in distinct populations and geographic contexts and include that information in a comprehensive conservation plan.

References

Allendorf FW, Leary RF, Spruell P, Wenburg JK (2001) The problems with hybrids: setting conservation guidelines. *Trends in Ecology & Evolution*, **16**, 613–622.

Anderson TM, Candille SI, Musiani M *et al.* (2009) Molecular and evolutionary history of melanism in North American gray wolves. *Science*, **323**, 1339–1343.

Ayres DR, Grotkopp E, Zaremba K *et al.* (2008) Hybridization between invasive *Spartina*

densiflora (Poaceae) and native *S. foliosa* in San Francisco Bay, California, USA. *American Journal of Botany*, **95**, 713–719.

Baird NA, Etter PD, Atwood TS *et al.* (2008) Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PloS One*, **3**, e3376.

Beatty GE, Philipp M, Provan J (2010) Unidirectional hybridization at a species' range boundary: implications for habitat tracking. *Diversity and Distributions*, **16**, 1–9.

Becker M, Gruenheit N, Steel M *et al.* (2013) Hybridization may facilitate in situ survival of endemic species through periods of climate change. *Nature Climate Change*, **3**, 1039.

Boughman JW (2001) Divergent sexual selection enhances reproductive isolation in sticklebacks. *Nature*, **411**, 944.

Bourret V, O'reilly PT, Carr JW, Berg PR, Bernatchez L (2011) Temporal change in genetic integrity suggests loss of local adaptation in a wild Atlantic salmon (*Salmo salar*) population following introgression by farmed escapees. *Heredity*, **106**, 500.

Brelsford A, Dufresnes C, Perrin N (2016) High-density sex-specific linkage maps of a European tree frog (*Hyla arborea*) identify the sex chromosome without information on offspring sex. *Heredity*, **116**, 177.

Brown JH (1984) On the relationship between abundance and distribution of species. *The American Naturalist*, **124**, 255–279.

Burgess KS, Morgan M, Deverno L, Husband BC (2005) Asymmetrical introgression between two Morus species (*M. alba*, *M. rubra*) that differ in abundance. *Molecular Ecology*, **14**, 3471–3483.

Butlin RK, Ritchie MG (2013) Pulling together or pulling apart: hybridization in theory and practice. *Journal of Evolutionary Biology*, **26**, 294–298.

Caeiro-Dias G, Luís C, Pinho C *et al.* (2018) Lack of congruence of genetic and niche divergence in *Podarcis hispanicus* complex. *Journal of Zoological Systematics and Evolutionary Research*, **56**, 479–492.

Capula M (1993) Natural hybridization in *Podarcis sicula* and *P. wagleriana* (Reptilia: Lacertidae). *Biochemical Systematics and Ecology*, **21**, 373–380.

Capula M (2002) Genetic evidence of natural hybridization between *Podarcis sicula* and *Podarcis tiliguerta* (Reptilia: Lacertidae). *Amphibia-Reptilia*, **23**, 313–321.

Carretero MA, Ribeiro R, Barbosa D, Sá-Sousa P, Harris DJ (2006) Spermatogenesis in two Iberian Podarcis lizards: relationships with male traits. *Animal Biology*, **56**, 1–

12.

Carretero MA, Sá-Sousa P, Barbosa D, Harris DJ, Pinho C (2002) Sintopía estricta entre *Podarcis bocagei* y *P. carbonelli*. *Boletín de la Asociación Herpetológica Española*, **13**, 20–24.

Catchen JM (2013) Stacks: an analysis tool set for population genomics. *Molecular Ecology*, **22**, 3124–3140.

Coyne JA, Orr HA (1997) "Patterns of speciation in *Drosophila*" revisited. *Evolution*, **51**, 295–303.

Danecek P, Auton A, Abecasis G *et al.* (2011) The variant call format and VCFtools. *Bioinformatics*, **27**, 2156–2158.

Dias G, Luis C, Pinho C, Kaliontzopoulou A (2016) A case of *Podarcis carbonelli* intake by *Podarcis virescens*. *Herpetology Notes*, **9**, 105–108.

Fitzpatrick BM (2012) Estimating ancestry and heterozygosity of hybrids using molecular markers. *BMC Evolutionary Biology*, **12**, 131.

Funk DJ, Nosil P, Etges WJ (2006) Ecological divergence exhibits consistently positive associations with reproductive isolation across disparate taxa. *Proceedings of the National Academy of Sciences of the United States of America*, **103**, 3209–3213.

Gompert Z, Alex Buerkle C (2010) INTROGRESS: a software package for mapping components of isolation in hybrids. *Molecular Ecology Resources*, **10**, 378–384.

Grossen C, Seneviratne SS, Croll D, Irwin DE (2016) Strong reproductive isolation and narrow genomic tracts of differentiation among three woodpecker species in secondary contact. *Molecular Ecology*, **25**, 4247–4266.

Hochkirch A, Lemke I (2011) Asymmetric mate choice, hybridization, and hybrid fitness in two sympatric grasshopper species. *Behavioral Ecology and Sociobiology*, **65**, 1637–1645.

Hoffmann AA, Sgrò CM (2011) Climate change and evolutionary adaptation. *Nature*, **470**, 479.

Hohenlohe PA, Amish SJ, Catchen JM, Allendorf FW, Luikart G (2011) Next-generation RAD sequencing identifies thousands of SNPs for assessing hybridization between rainbow and westslope cutthroat trout. *Molecular Ecology Resources*, **11**, 117–122.

Hohenlohe PA, Bassham S, Etter PD *et al.* (2010) Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. *PLoS Genetics*, **6**.

Irwin DE, Brelsford A, Toews DPL, MacDonald C, Phinney M (2009) Extensive

hybridization in a contact zone between MacGillivray's warblers *Oporornis tolmiei* and mourning warblers *O. philadelphia* detected using molecular and morphological analyses. *Journal of Avian Biology*, **40**, 539–552.

Jombart T (2008) adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics*, **24**, 1403–1405.

Jombart T, Ahmed I (2011) adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. *Bioinformatics*, **27**, 3070–3071.

Kaliontzopoulou A, Carretero MA, Llorente GA (2011a) Morphology of the *Podarcis* wall lizards (Squamata: Lacertidae) from the Iberian Peninsula and North Africa: patterns of variation in a putative cryptic species complex. *Zoological Journal of the Linnean Society*, **164**, 173–193.

Kaliontzopoulou A, Pinho C, Harris DJ, Carretero MA (2011b) When cryptic diversity blurs the picture: A cautionary tale from Iberian and North African *Podarcis* wall lizards. *Biological Journal of the Linnean Society*, **103**, 779–800.

Lepais O, Petit RJ, Guichoux E *et al.* (2009) Species relative abundance and direction of introgression in oaks. *Molecular Ecology*, **18**, 2228–2242.

Levin DA, Francisco-Ortega J, Jansen RK (1996) Hybridization and the extinction of rare plant species. *Conservation Biology*, **10**, 10–16.

Mayr E (1942) *Systematics and the origin of species, from the viewpoint of a zoologist*. Harvard University Press.

Milián-García Y, Ramos-Targarona R, Pérez-Fleitas E *et al.* (2015) Genetic evidence of hybridization between the critically endangered Cuban crocodile and the American crocodile: implications for population history and in situ/ex situ conservation. *Heredity*, **114**, 272.

Milne RI, Abbott RJ, Wolff K, Chamberlain DF (1999) Hybridization among sympatric species of Rhododendron (Ericaceae) in Turkey: morphological and molecular evidence. *American Journal of Botany*, **86**, 1776–1785.

Neaves LE, Zenger KR, Cooper DW, Eldridge MDB (2010) Molecular detection of hybridization between sympatric kangaroo species in south-eastern Australia. *Heredity*, **104**, 502.

Parchman TL, Gompert Z, Mudge J *et al.* (2012) Genome-wide association genetics of an adaptive trait in lodgepole pine. *Molecular Ecology*, **21**, 2991–3005.

Pereira RJ, Monahan WB, Wake DB (2011) Predictors for reproductive isolation in a ring

species complex following genetic and ecological divergence. *BMC Evolutionary Biology*, **11**, 194.

Pérez-Mellado V (1982) Algunos datos sobre la reproducción de dos especies de Lacertidae (Sauria, Reptilia) en el Sistema Central. *Bol. R. Soc. Esp. Hist. Nat. Secc. Biol*, **80**, 165–173.

Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE (2012) Double digest RADseq: an inexpensive method for *de novo* SNP discovery and genotyping in model and non-model species. *PloS One*, **7**, e37135.

Pinho C, Harris DJ, Ferrand N (2007) Comparing patterns of nuclear and mitochondrial divergence in a cryptic species complex the case of Iberian and North African wall lizards (Podarcis, Lacertidae). *Biological Journal of the Linnean Society*, **91**, 121–133.

Pinho C, Kaliontzopoulou A, Carretero MA, Harris DJ, Ferrand N (2009) Genetic admixture between the Iberian endemic lizards *Podarcis bocagei* and *Podarcis carbonelli*: Evidence for limited natural hybridization and a bimodal hybrid zone. *Journal of Zoological Systematics and Evolutionary Research*, **47**, 368–377.

Pinho C, Kaliontzopoulou A, Harris DJ, Ferrand N (2011) Recent evolutionary history of the Iberian endemic lizards *Podarcis bocagei* (Seoane, 1884) and *Podarcis carbonelli* Pérez-Mellado, 1981 (Squamata: Lacertidae) revealed by allozyme and microsatellite markers. *Zoological Journal of the Linnean Society*, **162**, 184–200.

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics*, **155**, 945–959.

Pujolar JM, Jacobsen MW, Als TD *et al.* (2014) Assessing patterns of hybridization between North Atlantic eels using diagnostic single-nucleotide polymorphisms. *Heredity*, **112**, 627.

Purcell J, Brelsford A, Wurm Y, Perrin N, Chapuisat M (2014) Convergent genetic architecture underlies social organization in ants. *Current Biology*, **24**, 2728–2732.

Renoult JP, Geniez P, Bacquet P, Benoit L, Crochet PA (2009) Morphology and nuclear markers reveal extensive mitochondrial introgressions in the Iberian Wall Lizard species complex. *Molecular Ecology*, **18**, 4298–4315.

Rhymer JM, Simberloff D (1996) Extinction by hybridization and introgression. *Annual review of ecology and systematics*, **27**, 83–109.

Rhymer JM, Williams MJ, Braun MJ (1994) Mitochondrial analysis of gene flow between New Zealand mallards (*Anas platyrhynchos*) and grey ducks (*A. superciliosa*). *The*

Auk, 970–978.

Rochette NC, Catchen JM (2017) Deriving genotypes from RAD-seq short-read data using Stacks. *Nature Protocols*, **12**, 2640.

Rubidge EM, Taylor EB (2005) An analysis of spatial and environmental factors influencing hybridization between native westslope cutthroat trout (*Oncorhynchus clarki lewisi*) and introduced rainbow trout (*O. mykiss*) in the upper Kootenay River drainage, British Columbia. *Conservation Genetics*, **6**, 369–384.

Sá-Sousa P (2001) Comparative chorology between *Podarcis bocagei* and *P . carbonellae* (Sauria: Lacertidae) in Portugal. *Revista Española de Herpetologia*, **15**, 85–97.

Sá-Sousa P, Harris DJ (2002) *Podarcis carbonelli* Perez-Mellado, 1981 is a distinct species. *Amphibia-Reptilia*, **23**, 459–468.

Sætre G (2013) Hybridization is important in evolution, but is speciation? *Journal of Evolutionary Biology*, **26**, 256–258.

Seehausen OLE, Takimoto G, Roy D, Jokela J (2008) Speciation reversal and biodiversity dynamics with hybridization in changing environments. *Molecular Ecology*, **17**, 30–44.

Senn H V, Pemberton JM (2009) Variable extent of hybridization between invasive sika (*Cervus nippon*) and native red deer (*C. elaphus*) in a small geographical area. *Molecular Ecology*, **18**, 862–876.

Sillero N, Argaña E, Matos C *et al.* (2012) Unexpected low population levels of *Podarcis carbonelli* in southern Salamanca (Spain). *Boletin de la Asociacion Herpetologica Espanola*, **23**, 69–73.

Sillero N, Carretero MA (2013) Modelling the past and future distribution of contracting species. The Iberian lizard *Podarcis carbonelli* (Squamata: Lacertidae) as a case study. *Zoologischer Anzeiger*, **252**, 289–298.

Sillero N, Leerschool T, Hollanders M, Beukema W (2014) No more *Podarcis carbonelli* in its type locality? *Herpetology Notes*, **7**, 751–754.

Steeves TE, Maloney RF, Hale ML, Tylianakis JM, Gemmell NJ (2010) Genetic analyses reveal hybridization but no hybrid swarm in one of the world's rarest birds. *Molecular Ecology*, **19**, 5090–5100.

Szymura JM, Barton NH (1986) Genetic analysis of a hybrid zone between the fire-bellied toads, *Bombina bombina* and *B. variegata*, near Cracow in southern Poland.

Evolution, **40**, 1141–1159.

Tarroso P, Pereira RJ, Martínez-Freiría F, Godinho R, Brito JC (2014) Hybridization at an ecotone: Ecological and genetic barriers between three Iberian vipers. *Molecular Ecology*, **23**, 1108–1123.

Tompkins DM, Mitchell RA, Bryant DM (2006) Hybridization increases measures of innate and cell-mediated immunity in an endangered bird species. *Journal of Animal Ecology*, **75**, 559–564.

vonHoldt BM, Brzeski KE, Wilcove DS, Rutledge LY (2018) Redefining the role of admixture and genomics in species conservation. *Conservation Letters*, **11**, e12371.

Vuillaume B, Valette V, Lepais O, Grandjean F, Breuil M (2015) Genetic evidence of hybridization between the endangered native species *Iguana delicatissima* and the invasive *Iguana iguana* (Reptilia, Iguanidae) in the Lesser Antilles: management implications. *PloS One*, **10**, e0127575.

Wayne RK, Shaffer HB (2016) Hybridization and endangered species protection in the molecular era. *Molecular Ecology*, **25**, 2680–2689.

Whitney KD, Randell RA, Rieseberg LH (2010) Adaptive introgression of abiotic tolerance traits in the sunflower *Helianthus annuus*. *New Phytologist*, **187**, 230–239.

Whittemore AT, Schaal BA (1991) Interspecific gene flow in sympatric oaks. *Proceedings of the National Academy of Sciences*, **88**, 2540–2544.

Wolf DE, Takebayashi N, Rieseberg LH (2001) Predicting the risk of extinction through hybridization. *Conservation Biology*, **15**, 1039–1053.

Yukilevich R (2012) Asymmetrical patterns of speciation uniquely support reinforcement in *Drosophila*. *Evolution*, **66**, 1430–1446.

Chapter 4. General Discussion

Lizards from the *P. hispanicus* complex became one of the most extensively studied lizard groups in the last 20 years, particularly from the point of view of phylogeny (Harris & Sá-Sousa 2001; Sá-Sousa & Harris 2002; Busack *et al.* 2005; Pinho *et al.* 2006; Lima *et al.* 2009; Kaliontzopoulou *et al.* 2011b), morphology (Kaliontzopoulou *et al.* 2007, 2011a, 2012) and ecophysiology (Carretero *et al.* 2006a; Veríssimo & Carretero 2009; Carneiro *et al.* 2015). One of the main outcomes of phylogenetic studies along these years was the uncovering of cryptic diversity (Lima *et al.* 2009; Kaliontzopoulou *et al.* 2011b) which has consequences for the other research fields. To date, 16 mitochondrial lineages were discovered, including eleven lineages in the Iberian Peninsula and five in North Africa (Kaliontzopoulou *et al.* 2011b). Eight lineages, identified on the basis of multilocus genetic data, morphology and ecology, are currently recognized as valid species or subspecies (Sá-Sousa 2001; Sá-Sousa & Harris 2002; Carbonell & Lizard 2003; Busack *et al.* 2005; Geniez *et al.* 2007, 2014), while others still lack taxonomic revision. The increasing number of genetic studies on this species complex and the detection of cryptic diversity raised the need of new studies accounting for this extensive diversity but also the opportunity to do comparative studies among distinct divergent lineages. Some aspects of the ecology of the newly uncovered taxa are still poorly known, like the distribution limits and ecological conditions that these species need to survive and reproduce. These aspects of the ecology are strictly connected with the possibility to establish contact zones and the potential for the occurrence of hybridization. Furthermore, this group offers an opportunity to compare mechanisms leading to reproductive isolation between distinct taxa with a range of divergences, how reproductive isolation evolved among the group and the consequences for species gene pool if hybridization occurs.

4.1. Distribution and co-occurrence of species from the *P. hispanicus* complex

One of the most important applications of ENMs is the identification of the species ecological niche (Barbosa *et al.* 2012). Together with georeferenced samples, ENM allows the development of robust models that relate biological diversity with sets of environmental variables in specific geographic contexts (Guisan & Zimmermann 2000; Sillero 2011; Warren 2012, 2013). Various studies across the Mediterranean Basin using

ENM have been carried out, aiming to describe species niche, to identify the potential distributions and the variables that determine them (e.g. Muñoz *et al.* 2005; Real *et al.* 2005; Martínez-Freiría *et al.* 2008; Beukema *et al.* 2010). These include studies on several species of the *P. hispanicus* complex (Sá-Sousa 2000; Román *et al.* 2006; Kaliontzopoulou *et al.* 2008; Sillero & Carretero 2013); however, all were restricted to one or few species and were limited in their geographic context. For this thesis the realized climatic niche (sensu Sillero 2011) was characterized on the basis of topographic and climatic variables of each lineage of the *P. hispanicus* species complex, identified which are the main variables influencing the niche and how they are related to the presence of each lineage, employing ENM techniques on genetically confirmed occurrence records across the distribution range of this group. Using the entire geographical region where the group is distributed takes into account the potential ecological range available for the each lineage. If the area is too small, part of the ranges of some variables may be left out of the analysis and lead to incorrect results (Barbosa *et al.* 2012). Additionally, if the study area is too small, the distribution of absences (or pseudo absences in our case) will be biased and so will the predicted ecological niche (Barbosa *et al.* 2012) .

We identified two main types of constraints for species distributions. Some climatic variables and the interspecific competition with other *Podarcis* species seem to be important factors shaping distributions among this group. Precipitation regimes and maximum temperature during the warmest season are highly important for the presence or absence of Iberian and North African wall lizards, which implies that generally these species do not occupy regions with extreme high temperatures and extreme low precipitation during the warmest months. This limitation of occupying arid, hot, and dry regions, like desert environments in North Africa, is most likely linked to the origin of the genus in more temperate Mediterranean environments (Harris *et al.* 2002; Sá-Sousa & Harris 2002; Lima *et al.* 2009). However, *Podarcis bocagei*, *P. g. lusitanicus* and *P. liolepis* occupy regions with higher precipitation levels during the warmest season in the studied area. For these three species, for *P. carbonelli* and for the Tunisia/Algeria group, the high values of annual precipitation are also important for their presence (Román *et al.* 2006; Kaliontzopoulou *et al.* 2008; Sillero & Carretero 2013).

According to ENMs, most of the lineages have at least some degree of realized climatic niche divergence. The discordance between niche and genetic divergence

suggests that differentiation was likely built in allopatric conditions, and parapatry or partial sympatry developed after secondary geographic contact between lineages. The partial niche divergence also suggests the potential for wide areas of sympatry between several lineages that are currently allopatric or parapatric. This implies that range limits in this species complex are explained by other factors than bioclimatic and topographic variables. Generally lineage distributions may be primarily limited by physiological tolerances (Kellermann *et al*., 2009), but a role for competition after secondary contact is suggested by comparing patterns of climatic realized niche and patterns of known distributions. Historical factors may also have a role, i.e., if one species is absent from an area due to competition and that area is suitable for both, it may be simply a consequence of one species colonized that area before the excluded species.

Such findings have direct implications for the establishment of contact zones. Partial niche divergence coupled with interspecific competition avoids the existence of extensive geographic regions of co-occurrence between pairs of lineages. Thus, ecological factors may act as partial mechanisms of geographical isolation.

4.2. Mechanisms of reproductive isolation

As discussed in the previous section, most pairs of *Podarcis* wall lizards have parapatric or allopatric distributions. Several pairs with parapatric distributions are known to meet in restricted and narrow contact zones and only few pairs are broadly sympatric. The analysis of the contact zone studied in this thesis revealed that a large fraction of individuals was assigned to one of the two parental species, clearly showing that the hybrid zones are bimodal and indicated the existence of strong reproductive isolation. For instance, we did not find any hybrids when *P. virescens* and *P. carbonelli* where found in strict syntopy. Without evident ecological or temporal barriers to gene flow between most of the co-occuring species, like for other *Podarcis* species (Pérez-Mellado 1982; Carretero *et al.* 2006b), intrinsic mechanisms, such as strong conspecific recognition, as it was observed between *P. bocagei* and *P. carbonelli* (Barbosa *et al.* 2005), may be acting as strong barriers to gene flow.

However, the detection of hybridization and introgression reveals that reproductive isolation is not complete between most pairs studied in more detail in this thesis (Chapter 2) but several barriers to interspecific gene flow are acting in the

maintenance of a strong reproductive isolation. We studied the mechanisms shaping the variation of reproductive isolation between *P. bocagei* and *P. carbonelli* across the genome. The geographic cline analysis shows a large concordance of loci cline centres with the area where both species co-occur, and most clines exhibit narrow widths that demonstrate the presence of barriers to gene flow, maintaining the genome cohesion of both species. This contrasts with the genomic cline analysis that evidences large heterogeneity of introgression patterns in the contact zone among loci. The discordance between geographic and genomic clines puts in evidence the existence of extrinsic barriers to gene flow. In the first study we identified only partial similarities between the climatic niches of these species. It is thus possible that environmental factors (including factors that remain to be identified) act as strong extrinsic barriers.

Almost half of loci exhibit increased or decreased patterns of introgression compared to genomic average and about 25% of the genomic regions analysed are potentially involved in intrinsic barriers to gene flow. In the absence of apparent extrinsic barriers, such intrinsic barriers seem strong enough to keep reproductive barriers that maintain species integrity. The alignment against *P. muralis* genome revealed that these regions are not gathered in restricted “genomic islands” but rather are well distributed across the genome. Nevertheless, the Z chromosome was found to have a distinct role in reproductive isolation relatively to the autosomes. Fewer Z-linked loci seem to be involved in stronger barriers to gene flow than autosomal chromosomes, suggesting a larger effect of the sexual chromosome in reproductive isolation in this system. Many studies have identified a distinct role for the X chromosome in reproductive isolation relatively to autosomal chromosomes (e.g. Orr 1987; Guénet *et al.* 1990; Storchova *et al.* 2004; Llopart 2012), which become a long-standing principle of speciation – the “large X effect” (Coyne & Orr 2004). Similarly, many studies recognized also the importance of the Z chromosome in reproductive isolation in species with ZW sex determining systems (e.g. Jiggins *et al.* 2001; Iyengar *et al.* 2002; Kirkpatrick & Hall 2004; Albert & Otto 2005; Storchová *et al.* 2010).

The patterns detected between *P. bocagei* and *P. carbonelli* in the late stages of speciation, as demonstrated by the steep geographic clines and the strong bimodality in the hybrid zone, were likely the result of distinct selective forces across the genome and across distinct stages of the evolution of reproductive isolation. It is possible that allopatric divergence might have had an important role in shaping such patterns but at

this stage we lost the ability to identify the genomic regions involved in the initial stages of reproductive isolation. Despite some regions do introgress, currently both extrinsic and intrinsic barriers to gene flow are acting together as strong reproductive isolation mechanisms between both species and maintaining the cohesion of both genomes.

A general outcome from these results is that the genetic basis of reproductive isolation in the late stages of speciation is complex and does not need the build-up of complete reproductive isolation across the genome. This complex arrangement of genomic regions involved in reproductive isolation may be a consequence of several interacting mechanisms and selective forces. Since most works in the study of the evolution of reproductive isolation focus on the early stages, this work intended to contribute to the study of the genomic architecture of reproductive isolation in the late stages of speciation, expecting that more similar studies in the future help to better understand the evolution of reproductive isolation.

4.3. Consequences of hybridization for *Podarcis carbonelli* persistence

We detected gene flow between *P. carbonelli* and most of its co-occurring species but globally genetic swamping does not seem to be a major overall threat to *P. carbonelli* since species cohesion is maintained. However, while hybridization and introgression seem to be relatively rare in some contact zones, interspecific gene flow may potentially be widespread in areas where *P. carbonelli* is sympatric with hybridizing species, for example *P. guadarramae lusitanicus* and *P. vaucheri* SSp. Moreover, the frequency of hybridization was found to be higher between *P. vaucheri* SSp. and *P. carbonelli* than in the other contact zones. This may be related with the low relative frequency of *P. carbonelli* in this area. This finding raises the question whether low *P. carbonelli*'s densities increase the rate of hybridization in other populations since *P. carbonelli* has several small fragmented populations across its distribution range with a continuous tendency for population decline. If so, hybridization may be context dependent and can hypothetically happen when one of the *Podarcis* species is much rarer than the other. Furthermore, with the unequivocal effects of global climate change together with other threatening factors, as population fragmentation, hybridization can affect the evolutionary processes by changing the way genes move around landscapes and by

introducing novel genotypes into populations (Hoffmann & Sgrò 2011).

P. carbonelli is one of the most threatened vertebrate species in the Iberian Peninsula. It is currently considered endangered by IUCN (Sá-Sousa et al. 2009) but no conservation plans with an integrated assessment of extinction risk were made to date; conservation measures directed at this species are almost inexistent. Issues like the consequences of interspecific mates and hybridization may be extremely important both when negative effects are an additional threat to populations (e.g. wasted reproductive effort in interspecific mates with other congeneric species can threat small and fragmented populations of *P. carbonelli*), or when positive effects may enhance fitness (e.g. introgression of new alleles in the genomic background of *P. carbonelli* positively selected in the new environmental conditions promoted by climate change). Thus the direct and indirect effects of hybridization as consequences of complex interactions with other threats emphasize their importance for management strategies. A comprehensive conservation plan for *P. carbonelli* is urgent, and the assessment and inclusion of the consequences of hybridization in distinct populations and geographic contexts cannot be neglected.

4.4. References

Albert AYK, Otto SP (2005) Sexual selection can resolve sex-linked sexual antagonism. *Science*, **310**, 119–121.

Barbosa D, Desfilis E, Carretero MA, Font E (2005) Chemical stimuli mediate species recognition in *Podarcis* wall lizards. *Amphibia-Reptilia*, **26**, 257–263.

Barbosa AM, Sillero N, Martínez-Freiría F, Real R (2012) Ecological niche models in Mediterranean herpetology: past, present, and future. *Ecological Modeling. Nova Science Publishers, Hauppauge, NY*, 173–204.

Beukema W, De Pous P, Donaire D *et al.* (2010) Biogeography and contemporary climatic differentiation among Moroccan Salamandra algira. *Biological Journal of the Linnean Society*, **101**, 626–641.

Busack S, Lawson R, Arjo W (2005) Mitochondrial DNA, allozymes, morphology and historical biogeography in the *Podarcis vaucheri* (Lacertidae) species complex. *Amphibia-Reptilia*, **26**, 239–256.

Carneiro D, García-Muñoz E, Kalionzopoulou A, Llorente GA, Carretero MA (2015)

Comparing ecophysiological traits in two *Podarcis* Wall lizards with overlapping ranges. *Salamandra*, **51**, 335–344.

Carretero MA, Marcos E, de Prado P (2006a) Intraspecific variation of preferred temperatures in the NE form of *Podarcis hispanica*. *Mainland and Insular Lacertid Lizards: a Mediterranean Perspective. Firenze University Press, Florence*, 55–64.

Carretero MA, Ribeiro R, Barbosa D, Sá-Sousa P, Harris DJ (2006b) Spermatogenesis in two Iberian *Podarcis* lizards: relationships with male traits. *Animal Biology*, **56**, 1–12.

Coyne JA, Orr HA (2004) Speciation. Sunderland, MA.

Geniez P, Cluchier A, Sá-Sousa P, Guillaume CP, Crochet PA (2007) Systematics of the *Podarcis hispanicus*-complex (Sauria, Lacertidae) I: Redefinition, morphology and distribution of the nominotypical taxon. *Herpetological Journal*, **17**, 69–80.

Geniez P, Sá-Sousa P, Guillaume CP, Cluchier A, Crochet PA (2014) Systematics of the *Podarcis hispanicus* complex (Sauria, Lacertidae) III: Valid nomina of the western and central Iberian forms. *Zootaxa*, **3794**, 1–51.

Guénet J-L, Nagamine C, Simon-Chazottes D, Montagutelli X, Bonhomme F (1990) Hst-3: an X-linked hybrid sterility gene. *Genetics Research*, **56**, 163–165.

Guisan A, Zimmermann NE (2000) Predictive habitat distribution models in ecology. *Ecological Modelling*, **135**, 147–186.

Harris DJ, Carranza S, Arnold EN, Pinho C, Ferrand N (2002) Complex biogeographical distribution of genetic variation within *Podarcis* wall lizards across the Strait of Gibraltar. *Journal of Biogeography*, **29**, 1257–1262.

Harris DJ, Sá-Sousa P (2001) Species distinction and relationships of the western Iberian *Podarcis* lizards (Reptilia, Lacertidae) based on morphology and mitochondrial DNA sequences. *Herpetological Journal*, **11**, 129–136.

Hoffmann AA, Sgrò CM (2011) Climate change and evolutionary adaptation. *Nature*, **470**, 479.

Iyengar VK, Reeve HK, Eisner T (2002) Paternal inheritance of a female moth's mating preference. *Nature*, **419**, 830.

Jiggins CD, Linares M, Naisbit RE *et al.* (2001) Sex-linked hybrid sterility in a butterfly. *Evolution*, **55**, 1631–1638.

Kaliontzopoulou A (2010) Proximate and evolutionary causes of phenotypic diversification: morphological variation in Iberian and North African Podarcis wall

lizards. *Departamento de Biología Animal*, **PhD thesis**, 227.

Kaliontzopoulou a., Brito JC, Carretero M a., Larbes S, Harris DJ (2008) Modelling the partially unknown distribution of wall lizards (*Podarcis*) in North Africa: ecological affinities, potential areas of occurrence, and methodological constraints. *Canadian Journal of Zoology*, **86**, 992–1001.

Kaliontzopoulou A, Carretero MA, Llorente GA (2007) Multivariate and geometric morphometrics in the analysis of sexual dimorphism variation in *Podarcis* lizards. *Journal of Morphology*, **268**, 152–165.

Kaliontzopoulou A, Carretero MA, Llorente GA (2011a) Morphology of the *Podarcis* wall lizards (Squamata: Lacertidae) from the Iberian Peninsula and North Africa: patterns of variation in a putative cryptic species complex. *Zoological Journal of the Linnean Society*, **164**, 173–193.

Kaliontzopoulou A, Carretero MA, Llorente GA (2012) Morphology of the *Podarcis* wall lizards (Squamata: Lacertidae) from the Iberian Peninsula and North Africa: Patterns of variation in a putative cryptic species complex. *Zoological Journal of the Linnean Society*, **164**, 173–193.

Kaliontzopoulou A, Pinho C, Harris DJ, Carretero MA (2011b) When cryptic diversity blurs the picture: A cautionary tale from Iberian and North African *Podarcis* wall lizards. *Biological Journal of the Linnean Society*, **103**, 779–800.

Kirkpatrick M, Hall DW (2004) Sexual selection and sex linkage. *Evolution*, **58**, 683–691.

Lima A, Pinho C, Larbes S *et al.* (2009) Relationships of *Podarcis* wall lizards from Algeria based on mtDNA data. *Amphibia-Reptilia*, **30**, 483–492.

Llopart A (2012) The rapid evolution of X-linked male-biased gene expression and the large-X effect in *Drosophila yakuba*, *D. santomea*, and their hybrids. *Molecular Biology and Evolution*, **29**, 3873–3886.

Martínez-Freiría F, Sillero N, Lizana M, Brito JC (2008) GIS-based niche models identify environmental correlates sustaining a contact zone between three species of European vipers. *Diversity and Distributions*, **14**, 452–461.

Muñoz AR, Real R, Barbosa AM, Vargas JM (2005) Modelling the distribution of Bonelli's eagle in Spain: implications for conservation planning. *Diversity and Distributions*, **11**, 477–486.

Orr HA (1987) Genetics of male and female sterility in hybrids of *Drosophila pseudoobscura* and *D. persimilis*. *Genetics*, **116**, 555–563.

Pérez-Mellado V (1982) Algunos datos sobre la reproducción de dos especies de Lacertidae (Sauria, Reptilia) en el Sistema Central. *Bol. R. Soc. Esp. Hist. Nat. Secc. Biol*, **80**, 165–173.

Pinho C, Ferrand N, Harris DJ (2006) Reexamination of the Iberian and North African *Podarcis* (Squamata: Lacertidae) phylogeny based on increased mitochondrial DNA sequencing. *Molecular Phylogenetics and Evolution*, **38**, 266–273.

Real R, Barbosa AM, Martínez-Solano Í, García-París M (2005) Distinguishing the distributions of two cryptic frogs (Anura: Discoglossidae) using molecular data and environmental modeling. *Canadian Journal of Zoology*, **83**, 536–545.

Román J, Ruiz G, Delibes M, Revilla E (2006) Factores ambientales condicionantes de la presencia de la lagartija de Carbonell *Podarcis carbonelli* (Pérez-Mellado, 1981) en la comarca de Doñana?. *Animal Biodiversity and Conservation*, **29**, 73–82.

Sá-Sousa P (2000) A predictive distribution model for the Iberian wall lizard (*Podarcis hispanicus*) in Portugal. *Herpetological Journal*, **10**, 1–11.

Sá-Sousa P (2001) Comparative chorology between *Podarcis bocagei* and *P. carbonellae* (Sauria : Lacertidae) in Portugal. *Revista Española de Herpetologia*, **15**, 85–97.

Sá-Sousa P, Harris DJ (2002) *Podarcis carbonelli* Perez-Mellado, 1981 is a distinct species. *Amphibia-Reptilia*, **23**, 459–468.

Sá-Sousa P, Pérez-Mellado V, Martínez-Solano I (2009) *Podarcis carbonelli*. *The IUCN Red List of Threatened Species* 2009: e.T61545A12512496. http://dx.doi.org/10.2305/IUCN.UK.2009.RLTS.T61545A12512496.en. Accessed on 23 September 2018.

Sillero N (2011) What does ecological modelling model? A proposed classification of ecological niche models based on their underlying methods. *Ecological Modelling*, **222**, 1343–1346.

Sillero N, Carretero MA (2013) Modelling the past and future distribution of contracting species. The Iberian lizard *Podarcis carbonelli* (Squamata: Lacertidae) as a case study. *Zoologischer Anzeiger*, **252**, 289–298.

Storchova R, Gregorová S, Buckiova D *et al.* (2004) Genetic analysis of X-linked hybrid sterility in the house mouse. *Mammalian Genome*, **15**, 515–524.

Storchová R, Reif J, Nachman MW (2010) Female heterogamety and speciation: Reduced introgression of the z chromosome between two species of nightingales.

Evolution, **64**, 456–471.

Veríssimo C V, Carretero MA (2009) Preferred temperatures of *Podarcis vaucheri* from Morocco: intraspecific variation and interspecific comparisons. *Amphibia-Reptilia*, **30**, 17–23.

Warren DL (2012) In defense of “niche modeling.” *Trends in Ecology and Evolution*, **27**, 497–500.

Warren DL (2013) “Niche modeling”: That uncomfortable sensation means it’s working. A reply to McInerny and Etienne. *Trends in Ecology and Evolution*, **28**, 193–194.

Chapter 5. Final Remarks and Future Research Directions

A general outcome from all the three studies presented in this thesis is that there are mechanisms preventing interspecific gene flow mediated by extrinsic and intrinsic factors across the *Podarcis hispanicus* complex. Frequently pre-mating mechanisms seem incomplete and hybridization occurs when distinct species meet in sympatric areas. However, several postzygotic mechanisms of selection prevent frequent hybridization and extensive introgression. Notably, even in the late stages os speciation, intrinsinc barriers to gene flow may not cover the entire genome but maintain genome cohesion.

These findings inspire the extension of the same kind of analysis to other contact zones. A comparative analysis across different contact zones between Iberian and North African *Podarcis* forms will be useful to define potential patterns of introgression across the group, to compare how intrinsic and extrinsic factors mechanisms act across different time scales and across pairs in distinct geographical contexts, ranging from parapatry in restricted areas to extensive sympatry. Given the patchy nature of the contact zones presented here and among other *Podarcis* species in the Iberian Peninsula and North Africa, it is not possible to establish a spatial axis from populations of one to other species. Genomic cline approaches offer the opportunity to investigate locus specific rates of introgression and compare patterns across the genome without the need for a distribution in a geographical gradient. Future work includes the use of such approaches across several contact zones.

Furthermore we already have data produced with RADseq for several sets of reference individuals covering their known distribution ranges both in the Iberian Peninsula and North Africa. These dataset are useful in other research fields. For instance, it allows to produce robust nuclear phylogenies, currently lacking if taking into account all known diversity. We can also perform evolutionary demographic studies, comparing different scenarios to test for historic gene flow and understand the role of gene flow in the early stages of speciation for the current patterns of diversity.

Appendix.

Appendix I. Lack of congruence of genetic and niche divergence in *Podarcis hispanicus* complex – Supporting Information

Appendix II. Genome wide patterns of interspecific admixture in a natural hybrid zone in late stages of speciation – Supporting Information

Appendix III. Evolution of sympatry without complete reproductive isolation: is hybridization relevant for *Podarcis carbonelli* conservation? – Supporting Information

Appendix IV. A case of *Podarcis carbonelli* intake by *Podarcis virescens*

Appendix I. Lack of congruence of genetic and niche divergence in *Podarcis hispanicus* complex - Supporting Information

Table S1.1. Location, mitochondrial identification and GenBank accession number for each sample corresponding to the presence points used for modelling.

Sample code	Lineage[1]	Locality	Country[2]	Latitude	Longitude	GenBank accession number	Locus[3]	Reference[4]
3.175	Pb	São Mamede do Coronado	PT	41.2853	-8.574517	HQ898069	*12S*	[11]
3.223	Pb	Subportela	PT	41.687433	-8.718117	HQ898071	*12S*	[11]
3.292	Pb	Palacios del Compludo	ES	42.45613	-6.45005	HQ898066	*12S*	[11]
3.296	Pb	Castro Laboreiro	PT	41.994817	-8.245917	HQ898061	*12S*	[11]
3.37	Pb	Moledo	PT	41.838567	-8.874067	EF081120	*ND4*	[14]
3.1383	Pb	Mindelo	PT	41.298483	-8.737017	HQ898064	*12S*	[11]
BEV.2056	Pb	Posadilla de la Vega, bridge over rio Tuerto (SW. of León)	ES	42.4002	-5.9518	KY461879	*D-loop*	[17]
BTA1	Pb	Tanes	ES	43.211167	-5.402533	DQ081064	*12S*	[13]
Cor1	Pb	A Coruña	ES	43.366667	-8.383333	EF081122	*ND4*	[14]
DB13169	Pb	Hospital de Orbigo	ES	42.46305	-5.877042	KY461838	*16S*	this study
DB3938	Pb	Tambre river	ES	42.978028	-8.4692	KY461839	*16S*	this study
DB4292	Pb	Torneros de la Valdería	ES	42.225422	-6.239401	HQ898073	*12S*	[11]
DB8035	Pb	Zimao	PT	41.45	-7.666667	EF081116	*ND4*	[14]
DB8086	Pb	Sanxenxo	ES	42.4	-8.816667	EF081116	*ND4*	[14]
DB8104	Pb	Sarria	ES	42.783333	-7.4	EF081117	*ND4*	[14]
DB8141	Pb	Taboadela	ES	42.233333	-7.816667	EF081128	*ND4*	[14]
DB8606	Pb	Caldas das Taipas	PT	41.466667	-8.333333	EF081126	*ND4*	[14]
DB8618	Pb	Valdovinho	ES	43.6	-8.116667	EF081122	*ND4*	[14]
DB8663	Pb	Penafiel	PT	41.2	-8.266667	EF081134	*ND4*	[14]
DB8665	Pb	Maia	PT	41.232862	-8.621576	HQ898063	*12S*	[11]
DB8760	Pb	Permedelos, Vila Verde	PT	41.738639	-8.423472	HQ898068	*12S*	[11]
DB9733	Pb	Espinho	PT	41.027433	-8.645533	EF081116	*ND4*	[14]
Gi30	Pb	Gião	PT	41.31295	-8.691633	HQ898004	*ND4*	[11]
Gpb6	Pb	Malpica	ES	43.316667	-8.800000	AF469425	*12S+*	[8]
MP3	Pb	Madalena	PT	41.103983	-8.661383	AF469423	*12S+*	[8]
Pbb1	Pb	Montesinho	PT	41.979267	-6.795317	AF372068	*COI+*	[7]
Pbb3	Pb	Vila Pouca de Aguiar	PT	41.445833	-7.672183	AF372075	*COI*	[7]
Pbb4	Pb	Serra do Geres	PT	41.718333	-8.166667	AF372070	*COI+*	[7]
Pbb5	Pb	Vairao	PT	41.330383	-8.6724	AF372087	*cytb+*	[7]
Pbb8	Pb	Braga	PT	41.535	-8.418333	AF372071	*COI+*	[7]
Pbb9	Pb	Viana do Castelo	PT	41.7	-8.818333	AF372076	*COI*	[7]
	Pb	Bridge over Rio Ulla, near Puentecesures	ES	42.73035	-8.636267	NA	*ND4*	[1]
	Pb	Mosteiro de la Armenteira	ES	42.464367	-8.7414	NA	*ND4*	[1]
	Pb	Praia de Louro	ES	42.75695	-9.104483	NA	*ND4*	[1]
	Pb	Punta Preguntoiro	ES	42.587000	-8.790750	NA	*ND4*	[1]
	Pb	Quarry near Caldas de Reis	ES	42.613383	-8.627817	NA	*ND4*	[1]

Sample code	Lineage[1]	Locality	Country[2]	Latitude	Longitude	GenBank accession number	Locus[3]	Reference[4]
	Pb	Esposende	PT	41.54675	-8.791483	AF133440	*12S+*	[6]
	Pb	Alvao	PT	41.35	-7.866667	EF081116	*ND4*	[14]
	Pb	Marco de Canavezes	PT	41	-8.069389	EF081132	*ND4*	[14]
	Pb	Between Mosteiro da la Armenteira and Mostero del Poio	ES	42.4609	-8.69875	NA	*ND4*	[1]
	Pb	Mirador de la Curota, near windmill plant of Serra de Barbanza	ES	42.648933	-8.9658	NA	*ND4*	[1]
4.159	Pc	El Acebuche	ES	37.04774	-6.565696	HQ898074	*12S*	[11]
Albc1	Pc	La Alberca	ES	40.466667	-6.083333	DQ081066	*12S+*	[13]
BEV.4061	Pc	first crossing towards Peña de Francia between El Cabaco and La Alberca	ES	40.543	-6.145	KY461880	*D-loop*	[17]
BEV.4074	Pc	San Joao de Ver, 3 km N of Santa Maria da Feira (S of Porto)	PT	40.951	-8.54	KY461881	*D-loop*	this study
BEV.4585	Pc	Doñana	ES	37.046	-6.567	KY461882	*D-loop*	this study
DB21502	Pc	Espinho	PT	41.027433	-8.645533	KY461837	*12S*	this study
DB21533	Pc	Santa Maria da Feira castle	PT	40.920982	-8.543151	KP455499	*16S+*	[5]
DB8202	Pc	S. Pedro de Moel	PT	39.75	-9.016667	EF081150	*ND4*	[14]
DB8205	Pc	Aveiro (Praia da Barra)	PT	40.631750	-8.746350	EF081152	*ND4*	[14]
DB8244	Pc	S. Pedro do Sul	PT	40.75	-8.066667	EF081154	*ND4*	[14]
DB8246	Pc	Tondela	PT	40.516667	-8.066667	EF081152	*ND4*	[14]
DB8261	Pc	Villasrubias	ES	40.316667	-6.616667	EF081139	*ND4*	[14]
DB8282	Pc	Carrico	PT	39.966667	-8.8	EF081147	*ND4*	[14]
DB8283	Pc	Satao	PT	40.716667	-7.716667	EF081153	*ND4*	[14]
DB8284	Pc	Meco	PT	38.466667	-9.166667	EF081146	*ND4*	[14]
DB8285	Pc	Cabo Raso	PT	38.7	-9.466667	EF081147	*ND4*	[14]
DB8303	Pc	Esmoriz	PT	40.616667	-8.75	EF081152	*ND4*	[14]
DB8320	Pc	Pendilhe	PT	40.883333	-7.816667	EF081152	*ND4*	[14]
DB8620	Pc	Sines	PT	37.972650	-8.869267	EF081145	*ND4*	[14]
DB9670	Pc	S. Jacinto	PT	40.662967	-8.748233	HQ898076	*12S*	[11]
Pcc1	Pc	Vale do Rossim, Serra da Estrela	PT	40.383333	-7.516667	AF372079	*cytb+*	[7]
Pcc2	Pc	Torreira	PT	40.763233	-8.710250	AF372080	*cytb+*	[7]
Pcc3	Pc	Monte Clérigo	PT	37.340588	-8.838951	AF372081	*cytb+*	[7]
Pcc4	Pc	Peniche	PT	39.350000	-9.368333	AF372056	*COI*	[7]
PR1	Pc	Playa de Rompeculos	ES	37.1	-6.75	AY214449	*12S+*	[9]
5.194	Pgg	Ciudad Rodrigo	ES	40.59295	-6.536333	HQ898107	*12S*	[11]
5.203	Pgg	Alba de Tormes	ES	40.82559	-5.51546	HQ898104	*12S*	[11]
BEV.2012	Pgg	El Tiemblo, in the village (SE. of Avila)	ES	40.415	-4.498	KY461877	*D-loop*	[17]
DB11155	Pgg	Cercedilla	ES	40.741041	-4.054232	KY461841	*16S*	this study

Sample code	Lineage[1]	Locality	Country[2]	Latitude	Longitude	GenBank accession number	Locus[3]	Reference[4]
DB8461	Pgg	Bejar	ES	40.383333	-5.766667	HQ898106	*12S*	[11]
DB8614	Pgg	Arévalo	ES	41.062071	-4.720288	HQ898105	*12S*	[11]
DB8615	Pgg	Las Ventas c/ Peña Aguilera	ES	39.616667	-4.216667	HQ898110	*12S*	[11]
DB8621	Pgg	El Piornal	ES	40.116667	-5.85	HQ898109	*12S*	[11]
DB8903	Pgg	Torrejon de la Calzada	ES	40.2	-3.8	HQ898111	*12S*	[11]
E210622	Pgg	Sierra de Gredos	ES	40.139945	-5.282891	AY132347	*cytb+*	[10]
Gual1	Pgg	Guadarrama	ES	40.683333	-4.083333	EU269560	*ND4*	[15]
HLA1	Pgg	La Alberca	ES	40.466667	-6.083333	EU269561	*ND4*	[15]
Maq1	Pgg	Maqueda	ES	40.066667	-4.366667	AY132319	*cytb+*	[10]
MNCN 11095	Pgg	Arenas de San Pedro	ES	40.2	-5.083333	AY234163	*cytb+*	[2]
MVZ 232040;	Pgg	San Martin del Pimpollar	ES	40.366667	-5.033333	AY234154	*cytb+*	[2]
Oro1	Pgg	Oropesa	ES	39.9199	-5.17465	AF469452	*12S+*	[8]
Trj1	Pgg	Trujillo	ES	39.460667	-5.8815	AF469450	*12S+*	[8]
Vil3	Pgg	Villacastin	ES	40.783333	-4.416667	EU269558	*ND4*	[15]
	Pgg	Peña de Francia	ES	40.543	-6.145	AY151906	*cytb+*	[3]
5.225	Pgl	Tudera	ES	41.41689	-6.21043	HQ898099	*12S*	[11]
5.23	Pgl	Moledo	PT	41.838567	-8.874067	KY461891	*ND4*	this study
5.247	Pgl	Ledesma	ES	41.09175	-5.9979	HQ898086	*12S*	[11]
5.262	Pgl	Sta. Eulalia	ES	42.03222	-6.2683	HQ898097	*12S*	[11]
5.143	Pgl	Alvao NP	PT	41.349683	-7.792417	HQ898077	*12S*	[11]
Anc2	Pgl	Los Ancares	ES	42.669633	-6.726967	EU269553	*ND4*	[15]
And3	Pgl	Vale do Rossim, Serra da Estrela	PT	40.383333	-7.516667	HQ898054	*16S*	[11]
BEV.2052	Pgl	1 km past Ocero on the road to El Espino (N. of Ponferrada)	ES	42.706	-6.63	KY461878	*D-loop*	this study
BEV.8332	Pgl	8 km past Rihonor de Castilla on road to Varge (Serra de Montesinho)	PT	41.91133	-6.63945	KY461885	*D-loop*	this study
BEV.8334	Pgl	1 km past Ungilbe on road to Puebla de Sanabria	ES	42.03764	-6.61951	KY461886	*D-loop*	this study
BEV.8346	Pgl	crossing of road to Brañuela along road N-VI, 900 m. NE of La Silva	ES	42.6087	-6.2588	KY461887	*D-loop*	this study
DB11711	Pgl	Allende	ES	43.215987	-4.596735	KY461840	*16S*	this study
DB16699	Pgl	Sabugal	PT	40.348877	-7.092634	KY461888	*ND4*	this study
DB1730	Pgl	Fornillos (de Aliste)	ES	41.656341	-6.192236	HQ898084	*12S*	[11]
DB1734	Pgl	Crestuma Castle	PT	41.06644	-8.504137	HQ898083	*12S*	[11]
DB1751	Pgl	Near Sta. Eulalia	ES	42.03736	-6.26429	HQ898090	*12S*	[11]
DB1753	Pgl	Rio Casares	ES	42.926643	-5.772748	HQ898092	*12S*	[11]
DB1758	Pgl	Rio Negro, Peque	ES	42.04394	-6.2648	HQ898094	*12S*	[11]
DB8322	Pgl	Geres	PT	41.718333	-8.166667	HQ898085	*12S*	[11]
DB8398	Pgl	Chelos, Gaia	PT	41.052478	-8.484142	HQ898081	*12S*	[11]
DB8399	Pgl	Oliveira do Hospital	PT	40.408722	-7.924528	HQ898005	*ND4*	[11]

Sample code	Lineage[1]	Locality	Country[2]	Latitude	Longitude	GenBank accession number	Locus[3]	Reference[4]
DB8400	Pgl	Sobreira (Chaves)	PT	41.75236	-7.37592	HQ898096	*12S*	[11]
DB8401	Pgl	Vila Chã (Vale de Cambra)	PT	41.85372	-8.402	HQ898101	*12S*	[11]
DB8403	Pgl	Vinhais	PT	41.834083	-7.003502	HQ898102	*12S*	[11]
DB8409	Pgl	Celanova	ES	42.15	-7.966667	HQ898079	*12S*	[11]
DB8411	Pgl	Murça	PT	41.405321	-7.453975	HQ898089	*12S*	[11]
DB8416	Pgl	Zamora	ES	41.5	-5.75	HQ898103	*12S*	[11]
DB8609	Pgl	Cidadelhe	PT	40.909803	-7.109462	HQ898082	*12S*	[11]
DB8612	Pgl	UTM NG23 aka "Serra d'Arga"	PT	41.84986	-8.71	HQ898095	*12S*	[11]
DB8653	Pgl	Barrocal do Douro	PT	41.379018	-6.351142	HQ898078	*12S*	[11]
DB8669	Pgl	Lourosa	PT	41.220972	-8.069389	HQ898088	*12S*	[11]
DB8671	Pgl	Chavães	PT	41.182444	-8.020389	HQ898080	*12S*	[11]
DB8672	Pgl	Sto. Estevao	PT	41.762278	-7.402389	HQ898098	*12S*	[11]
FT12	Pgl	Tua	PT	41.218333	-7.368333	EU269554	*ND4*	[15]
Pen2	Pgl	Pendilhe	PT	40.883333	-7.816667	EU269556	*ND4*	[15]
Ph1	Pgl	Vila Real	PT	41.285	-7.75	AF372080	*cytb+*	[7]
Ph2	Pgl	Montesinho	PT	41.979267	-6.795317	AF372061	*COI+*	[7]
Rua1	Pgl	Vila de Rua	PT	40.95	-7.566667	AF469444	*12S+*	[8]
	Pgl	Between Mosteiro da la Armenteira and Mostero del Poio	ES	42.4609	-8.69875	NA	*ND4*	[1]
	Pgl	Mirador de la Curota, near windmill plant of Serra de Barbanza	ES	42.648933	-8.9658	NA	*ND4*	[1]
	Pgl	Ardia	ES	42.458783	-8.879617	NA	*ND4*	[1]
	Pgl	Castineiras, slope of moun	ES	42.54335	-8.996133	NA	*ND4*	[1]
	Pgl	Mirador, Con de la Siradella, Peninsula del Grove	ES	42.479417	-8.860333	NA	*ND4*	[1]
	Pgl	Playa de Lanzada	ES	42.431	-8.87315	NA	*ND4*	[1]
	Pgl	Punta Moreiras	ES	42.485233	-8.890350	NA	*ND4*	[1]
	Pgl	Punta San Vicente	ES	42.456200	-8.925350	NA	*ND4*	[1]
9.76	PhAM	Canada del Provencio	ES	38.518033	-2.35315	HQ898185	*12S*	[11]
9.79	PhAM	Sierra de la Oliva	ES	38.764883	-0.982583	HQ898190	*12S*	[11]
DB1285	PhAM	Sierra de la Pila	ES	38.264363	-1.18982	HQ898192	*12S*	[11]
DB1817	PhAM	Camino del Tobalejo	ES	38.65	-2.22	HQ898184	*12S*	[11]
DB1841	PhAM	El Pardal	ES	38.485309	-2.287438	HQ898187	*12S*	[11]
DB1878	PhAM	Montealegre del Castillo	ES	38.830755	-1.339061	HQ898188	*12S*	[11]
DB3861	PhAM	Sierra de Callosa del Segura	ES	38.120213	-0.906044	HQ898189	*12S*	[11]
9.1	PhG	Sierra de Espuña	ES	37.81985	-1.582683	HQ898181	*12S*	[11]
9.6	PhG	Cartagena	ES	37.603283	-1.007517	HQ898174	*12S*	[11]
9.68	PhG	Caravaca de la Cruz	ES	38.10745	-1.85905	HQ898173	*12S*	[11]
BEV.7050	PhG	4 km past Berja on road to Dalía	ES	36.836917	-2.90916	FJ208706	*D-loop*	[18]

Sample code	Lineage[1]	Locality	Country[2]	Latitude	Longitude	GenBank accession number	Locus[3]	Reference[4]
BEV.7330	PhG	bridge over rio Mula at E. of Albudeite	ES	38.0278	-1.3819	FJ208730	*D-loop*	[18]
BEV.7331	PhG	Bullas, E end of Av. de Murcia	ES	38.0519	-1.6515	FJ208731		[18]
BEV.7344	PhG	Rambla de Chirivel, 1 km E. of Vélez Rubio	ES	37.6447	-2.0592	FJ208714	*D-loop*	[18]
BEV.7346	PhG	castel of Vélez Blanco (Sierra de Maria)	ES	37.6903	-2.0987	FJ208715	*D-loop*	[18]
BEV.7353	PhG	7 km past Puebla de Don Fadrique on road to Maria	ES	37.9128	-2.4001	EU269580	*ND4+*	[15]
BEV.7369	PhG	2 km N. of Tiscar	ES	37.7739	-3.0227	FJ208774	*D-loop*	[18]
BEV.7382	PhG	dam of the Embalse del Negratín (NW. Baza)	ES	37.5607	-2.9572	FJ208713	*D-loop*	[18]
DB2961	PhG	Rio Castril river source	ES	37.908395	-2.749222	HQ898180	*12S*	[11]
DB3841	PhG	Embalse de La Pedrera	ES	38.006363	-0.857593	HQ898176	*12S*	[11]
DB3851	PhG	Rambla del Cañar-Cartagena		37.605424	-1.164214	HQ898179	*12S*	[11]
DB8647	PhG	Láujar de Andarax	ES	36.993834	-2.889226	HQ898178	*12S*	[11]
Gal3	PhG	Galera	ES	37.741783	-2.549317	DQ081070	*12S+*	[13]
DB1008	JS	Âit Hani	MA	31.80184	-5.46698	KY461848	*16S*	this study
DB1022	JS	Agoudal	MA	32.03531	-5.46745	KY461847	*16S*	this study
DB11031	JS	Tizi-n'-Melloul	MA	30.8081	-7.58367	HQ898009	*ND4*	[11]
DB11253	JS	W of Tachakoucht	MA	30.8	-7.55	KY461846	*16S*	this study
DB11266	JS	N of Tachakoucht	MA	30.805617	-7.543783	KY461845	*16S*	this study
DB1663	JS	Road to Jbel Siroua	MA	30.788062	-7.593582	HQ898183	*12S*	[11]
DB980	JS	Âit-Taddert	MA	32.126503	-5.304087	KY461849	*16S*	this study
Js1	JS	Jbel Siroua	MA	30.746966	-7.609283	AY132315	*cytb+*	[10]
PH184	JS	5Km SE Jbel Siroua	MA	30.712385	-7.636617	EU269585		[15]
Aza879	TAg[b]	Azazga	ALG	36.753433	4.424833	GQ856131	*12S+*	[12]
DjeA31	TAg[c]	Djebel Aures	ALG	35.350117	6.621867	GQ856125	*12S+*	[12]
Ham1	TAg[c]	Hamla	ALG	35.5797	6.07625	GQ856127	*12S+*	[12]
8.48	TAg[a]	Feidja NP	TN	36.504433	8.313717	HQ898013	*ND4*	[11]
8.553	TAg[a]	Cap Negro	TN	37.068566	9.04645	HQ898012	*ND4*	[11]
DB1595	TAg[a]	Jbel Goraa	TN	36.490671	9.151754	HQ898221	*12S*	[11]
E30051	TAg[a]	N Ain Draham	TN	36.72105	8.67748	AY132338	*cytb+*	[10]
E30052	TAg[a]	S Ain Draham	TN	36.72105	8.677483	AY132339	*cytb+*	[10]
E30056	TAg[a]	10Km S Tabarca	TN	36.864478	8.726159	AY132342	*cytb+*	[10]
E30057	TAg[a]	Ain Draham	TN	36.772975	8.685613	AY132343	*cytb+*	[10]
Edo33	TAg[a]	Edough	ALG	36.883333	7.616667	GQ856129	*12S+*	[12]
Elk32	TAg[a]	El Kala	ALG	36.834439	8.415283	GQ856130	*12S+*	[12]
Jug2	TAg[a]	Jughourta Table	TN	35.732	8.40185	GQ856133	*12S+*	[12]
LK6	TAg[a]	Le Kef	TN	36.18468	8.7103	DQ081071	*12S+*	[13]
OK1	TAg[a]	Oued Kebir	TN	36.777	8.69593	DQ081072	*12S+*	[13]
9.22	V	Boniche	ES	39.985383	-1.628717	HQ898197	*12S*	[11]

Sample code	Lineage[1]	Locality	Country[2]	Latitude	Longitude	GenBank accession number	Locus[3]	Reference[4]
9.28	V	Xativa	ES	38.98575	-0.520017	HQ898219	*12S*	[11]
9.8	V	Castillo de la Calahora	ES	37.183467	-3.065133	HQ898200	*12S*	[11]
10.45	V	Cazorla, Nava de San Pedro	ES	37.891867	-2.864717	HQ898201	*12S*	[11]
B2	V	La Casella, Alzira	ES	39.12511	-0.386231	HQ898210	*12S*	[11]
B3	V	St Esperit, Gilet Sagunt	ES	39.669866	-0.348242	HQ898216	*12S*	[11]
BEV.1875	V	Calomarde, 17 km past Albarracín on the road towards Frias de Albarracín	ES	40.372	-1.574	KY461895	*ND4*	this study
BEV.3916	V	Alcudia de Crespins	ES	38.9751	-0.5887	FJ208743	*D-loop*	[18]
BEV.3919	V	Gandia	ES	38.9679	-0.1746	FJ208742	*D-loop*	[18]
BEV.3929	V	3 km E. of Betera	ES	39.5720	-0.4823	FJ208750	*D-loop*	[18]
BEV.3932	V	5,5 km S. of Nules	ES	39.8245	-0.2103	FJ208752	*D-loop*	[18]
BEV.4545	V	Uña (Serrania de Cuenca)	ES	40.224	-1.978	FJ208757	*D-loop*	[18]
BEV.4705	V	Camino del Cabañal, Valencia city	ES	39.4745	-0.3393	FJ208747	*D-loop*	[18]
BEV.4726	V	Ribarroja del Turia	ES	39.55	-0.57	FJ208749	*D-loop*	[18]
BEV.4733	V	Albufera, Muntanyeta dels Sants, near Sueca	ES	39.2421	-0.3163	FJ208745	*D-loop*	[18]
BEV.7017	V	3 km past Buena Vista on road CV.746 to Calpe	ES	38.66557	0.08871	FJ208733	*D-loop*	[18]
BEV.7264	V	W.NW. entrance of Alquerias del Niño Perdido (W.NW. of Burriana)	ES	39.8993	-0.1159	FJ208755	*D-loop*	[18]
BEV.7271	V	W. entrance of Grao de Burriana	ES	39.8688	-0.0732	FJ208754	*D-loop*	[18]
BEV.7279	V	bridge over rio Albaida, 500m W. of Manuel (N. of Xátiva)	ES	39.05105	-0.50007	FJ208744	*D-loop*	[18]
BEV.7280	V	bridge over rio Clariano, 4 km NE. of Ontinyent (S.SW. Xátiva)	ES	38.85465	-0.5853	FJ208740	*D-loop*	[18]
BEV.7294	V	road CV 50 at Catadau (S.SW. of Valencia)	ES	39.27425	-0.5765	FJ208746	*D-loop*	[18]
BEV.7309	V	16 km past Bocairent on road to Ibi (= 2 km NNW of El Baradello Gelat)	ES	38.7179	-0.5299	FJ208735	*D-loop*	[18]
BEV.7310	V	chapel 1 km N. of Ibi on road CV. 801	ES	38.6344	-0.5761	FJ208732	*D-loop*	[18]
BEV.7333	V	8 km past Sta Maria de Nieva on road A.327 to Puerto de Sta Maria de Nieva	ES	37.5295	-2.0041	KY461834	*12S*	this study
BEV.7337	V	5 km past Sta Maria de Nieva on road A.327 to Puerto de Sta Maria de Nieva	ES	37.4982	-1.9887	FJ208711	*D-loop*	[18]
BEV.7343	V	6 km past Sta Maria de Nieva on road A.327 to Puerto de Sta Maria de Nieva	ES	37.5512	-1.9962	FJ208712	*D-loop*	[18]

Sample code	Lineage[1]	Locality	Country[2]	Latitude	Longitude	GenBank accession number	Locus[3]	Reference[4]
BEV.7357	V	Sierra de la Hoya del Espino, 5 km past "El Puerto del Pinar" on road to Santiago de la Espada [38,0700°N/2,5560 °W/1550 m]	ES	38.07	-2.556	KY461898	*ND4*	this study
BEV.7376	V	4 km past Burunchel on road to Puente de las Herrerias (Sierra de Cazorla)	ES	37.9509	-2.9398	KY461899	*ND4*	this study
BEV.7400	V	dam of the Embalse del Negratín (NW. Baza)	ES	38.7179	-0.5299	FJ208736	*D-loop*	[18]
BEV.9851	V	Orange plantations near Sagunto	ES	39.7054	-0.2598	KY461912	*ND4*	this study
BEV.9854	V	orange plantations near Rafelbuñol (N. of Valencia)	ES	39.6144	-0.3502	KY461913	*ND4*	this study
BEV.9855	V	between Maises and the rio Turia	ES	39.5043	-0.4771	KY461914	*ND4*	this study
BEV.9856	V	Massanassa, in the village, Carrer del Palleter (S. of Valencia)	ES	39.4112	-0.3954	KY461915	*ND4*	this study
BEV.9858	V	1 km N. of Mareny Blau	ES	39.2262	-0.2574	KY461916	*ND4*	this study
BEV.9862	V	Mareny de San Llorenc de Cullera (S.of Valencia)	ES	39.2195	-0.2505	KY461917	*ND4*	this study
BEV.9873	V	southern exit of Ibi on road to Tibi	ES	38.6119	-0.5797	KY461918	*ND4*	this study
BEV.9874	V	Ibi town center, upper parts	ES	38.6255	-0.5711	KY461919	*ND4*	this study
BEV.9879	V	7 km past Sella on road to Alcolecha	ES	38.6372	-0.3166	KY461920	*ND4*	this study
BEV.9881	V	1 km past Port de Tudons on road to Orxeta	ES	38.6444	-0.3194	KY461921	*ND4*	this study
BEV.9884	V	50 m past the Port de Tudons on road to Alcolecha	ES	38.6509	-0.3246	KY461922	*ND4*	this study
BEV.9885	V	1 km past Alcolecha on road to Orxeta	ES	38.6716	-0.3308	KY461923	*ND4*	this study
BEV.9889	V	upper parts of Benifato village	ES	38.6738	-0.2303	KY461924	*ND4*	this study
BEV.9891	V	Castello de Guadalest	ES	38.676	-0.1991	KY461925	*ND4*	this study
BEV.9895	V	4 km past Alcoy going to Benasau	ES	38.6925	-0.4433	KY461926	*ND4*	this study
BEV.9896	V	2 km past Alcoy going to Benasau	ES	38.6865	-0.4587	KY461927	*ND4*	this study
BEV.9903	V	exit of Banyeres on road to Villena	ES	38.7196	-0.6662	KY461928	*ND4*	this study
CU	V	Ciudad Encantada	ES	40.207397	-2.005227	HQ898203	*12S*	[11]
Cue1	V	Cuenca	ES	40.066667	-2.133333	AF469430	*cybt+*	[8]
DB13475	V	La Virgen de la Vega, Gúdar	ES	40.369282	-0.670606	KY461850	*16S*	this study
DB1735	V	Guadalquivir river source	ES	37.839145	-2.974465	HQ898208	*12S*	[11]
DB1791	V	500m from Los Negros Camping	ES	38.271465	-2.6061	HQ898194	*12S*	[11]
DB1834	V	Calar de Mundo	ES	38.445081	-2.411088	HQ898199	*12S*	[11]
DB1853	V	Rio Madera	ES	38.245624	-2.628203	HQ898214	*12S*	[11]
DB1879	V	Barranco de Guadalentin	ES	37.871042	-2.87659	HQ898195	*12S*	[11]

Sample code	Lineage[1]	Locality	Country[2]	Latitude	Longitude	GenBank accession number	Locus[3]	Reference[4]
DB1887	V	Fte la Reina	ES	37.94373	-2.832242	HQ898205	*12S*	[11]
DB1895	V	Venta Benito	ES	38.173823	-2.600902	HQ898218	*12S*	[11]
DB1898	V	Fuente de la Garganta	ES	37.897253	-2.893832	HQ898207	*12S*	[11]
DB3053	V	Pico Cabañas	ES	37.814097	-2.9546	HQ898211	*12S*	[11]
DB3857	V	Revolcadores	ES	38.076936	-2.284371	HQ898213	*12S*	[11]
DB3858	V	Sierra Espuña-Zona Pozos de Nieve	ES	37.869971	-1.571928	HQ898215	*12S*	[11]
DB8628	V	Puebla del Salvador	ES	39.566667	-1.666667	HQ898212	*12S*	[11]
DB8630	V	Cortijo Becerra, Guadix	ES	37.3	-3.133333	HQ898204	*12S*	[11]
DB8646	V	Bunyol	ES	39.418258	-0.790769	HQ898198	*12S*	[11]
Mot1	V	Motilla del Palancar	ES	39.566667	-1.883333	AY132320	*cytb+*	[10]
Prg1	V	Puerto de la Ragua	ES	37.112467	-3.033117	AY132348	*cytb+*	[10]
	V	Burjasot	ES	39.516667	-0.416667	AJ224406	*cytb*	[4]
	V	Castellon	ES	39.983333	-0.033333	AF052634	*cytb*	[4]
	V	El Grao	ES	39.983642	0.020849	AJ004904	*cytb*	[4]
	V	El Saler	ES	39.383333	-0.333333	AJ004909	*cytb*	[4]
1.113	PI	Rio Segre	ES	42.368967	1.759917	HQ898171	*12S*	[11]
1.15	PI	Calomarde	ES	40.371917	-1.577733	HQ898166	*12S*	[11]
Bar1	PI	Barcelona	ES	41.383333	2.183333	AF469432	*cybt+*	[8]
BEV.10936	PI	service area of "Village Catalan", along A9 highway SSE of Villemolaque	FR	42.578	2.8483	KY461863	*16S*	this study
BEV.11057	PI	Mas Barou, 2 km N.N.NE. of Céret	FR	42.5031	2.7541	KY461864	*16S*	this study
BEV.1253	PI	Valras-Plage	FR	43.246987	3.292936	KY461858	*16S*	this study
BEV.1455	PI	Mèze, town center	FR	43.426324	3.606538	KY461859	*16S*	this study
BEV.1801	PI	Villanueva de Huerva, in the village	ES	41.352	-1.036	FJ208765	*D-loop*	[18]
BEV.1807	PI	Aguarón, in the village (province of Zaragoza)	ES	41.339	-1.269	FJ208764	*D-loop*	[18]
BEV.1808	PI	Codos (between Belchite and Cariñena)	ES	41.293	-1.374	KY461893	*ND4*	this study
BEV.1865	PI	Albarracín, in the city	ES	40.408	-1.444	KY461894	*ND4*	this study
BEV.2125	PI	Almenar de Soria	ES	41.6823	-2.2000	FJ208770	*D-loop*	[18]
BEV.3942	PI	6 km past Vinaros on road to Amposta, at the province border	ES	40.5259	0.5061	FJ208761	*D-loop*	[18]
BEV.4448	PI	Road D.117 at La Jonquière, Quillan	FR	42.878996	2.172717	KY461835	*12S*	this study
BEV.4550	PI	Béhobie	FR	43.3432	-1.7521	KY461836	*12S*	this study
BEV.4551	PI	2 km N. of Farena (N. of Reus)	ES	41.33	1.08	FJ208763	*D-loop*	[18]
BEV.4614	PI	service area of "Village Catalan", along A9 highway SSE of Villemolaque	FR	42.577433	2.84728	KY461860	*16S*	this study

Sample code	Lineage[1]	Locality	Country[2]	Latitude	Longitude	GenBank accession number	Locus[3]	Reference[4]
BEV.8356	PI	crossing of road to Malleval 300 m S. of the village	FR	45.38178	4.726956	KY461861	*16S*	this study
BEV.8357	PI	road D329, 800m SSE. of "Cap de Côte" (= 2,3 km NE. of Mandagou)	FR	44.04223	3.61138	FJ208772	*D-loop*	[18]
BEV.8361	PI	800m past "Col des Vielles" on road D323 to Taleyrac (= 2 km N. of Mandagou)	FR	44.04636	3.63358	KY461862	*16S*	this study
BEV.9815	PI	SE. edge of San Miquel de Montroig	ES	41.0311	0.9666	KY461900	*ND4*	this study
BEV.9818	PI	N.E. exit of San Miquel de Montroig on road to Cambrils	ES	41.0337	0.968	KY461901	*ND4*	this study
BEV.9823	PI	Just N of San Carles de la Rapita harbour	ES	40.6212	0.5985	KY461902	*ND4*	this study
BEV.9826	PI	6 km. past San Carles de la Rapita on road to Amposta	ES	40.6716	0.5887	KY461903	*ND4*	this study
BEV.9830	PI	eastern exit of Amposta on the southern bank of the Ebro river	ES	40.7093	0.5877	KY461904	*ND4*	this study
BEV.9833	PI	L'Aldea (SE. of Tortosa)	ES	40.7417	0.6122	KY461905	*ND4*	this study
BEV.9836	PI	3 km past Peñiscola on road to AP-7 highway	ES	40.3911	0.3759	KY461906	*ND4*	this study
BEV.9838	PI	orange cultivations near Torreblanca	ES	40.2237	0.2062	KY461907	*ND4*	this study
BEV.9843	PI	Calle San Isidro at N. exit of Oropesa del Mar	ES	40.0962	0.1346	KY461908	*ND4*	this study
BEV.9844	PI	just north of N. de Benicasim	ES	40.0566	0.0883	KY461909	*ND4*	this study
BEV.9845	PI	N. entrance of Benicasim	ES	40.0581	0.0856	KY461910	*ND4*	this study
BEV.9847	PI	W. suburbs of Castellón de la Plana, below the Magdalena hospital	ES	40.0031	-0.0709	KY461911	*ND4*	this study
DB13429	PI	S of Villaescusa de las Torres	ES	42.755393	-4.255153	KY461854	*16S*	this study
DB13470	PI	Barranc de l'Avellanar, Penyagolosa	ES	40.249969	-0.335148	KY461852	*16S*	this study
DB13474	PI	La Virgen de la Vega, Gúdar	ES	40.369282	-0.670606	KY461853	*16S*	this study
DB14667	PI	Sante Fe del Montseny	ES	41.76	2.47	KY461855	*16S*	this study
DB1716	PI	Aiguamolls del Emporda	ES	42.222	3.09234	HQ898163	*12S*	[11]
DB1731	PI	Monasterio de Moreruela	ES	41.81186	-5.77805	HQ898169	*12S*	[11]
DB1732	PI	Torredembarra	ES	41.151750	1.431183	HQ898172	*12S*	[11]
DB1762	PI	Les Solans		41.708984	1.98924	HQ898168	*12S*	[11]
DB5140	PI	Santa Barbara	ES	40.7113	0.50801	KY461856	*16S*	this study
DB5887	PI	Oriñon	ES	43.401027	-3.326262	KY461851	*16S*	this study
DB8605	PI	Near Sopeira	ES	42.309327	0.740439	HQ898170	*12S*	[11]
DB8613	PI	Castrillo de la Vega	ES	41.651982	-3.781494	HQ898167	*12S*	[11]
DB8664	PI	Bordils		42.045433	2.909819	HQ898165	*12S*	[11]

Sample code	Lineage[1]	Locality	Country[2]	Latitude	Longitude	GenBank accession number	Locus[3]	Reference[4]
DB9604	Pl	Alcolea del Pinar	ES	41.033333	-2.466667	HQ898164	*12S*	[11]
DB9935	Pl	Zaragoza	ES	41.652839	-0.894501	KY461857	*16S*	this study
E210622	Pl	AD	AD	42.5	1.516667	AY132347	*cytb+*	[10]
Get1	Pl	Getaria	ES	43.300000	-2.200000	EU269574	*ND4*	[15]
Gir1	Pl	Girona	ES	41.983333	2.816667	AF469440	*cybt+*	[8]
Med1	Pl	Medinaceli	ES	41.16225	-2.417767	AF469436	*cybt+*	[8]
MNCN 11094	Pl	Mayorga	ES	42.166667	-5.266667	AY234160	*cytb+*	[2]
	Pl	Cago Higuer (Jaizkibel)	ES	43.388821	-1.794648	NA	*12S+*	[19]
	Pl	Monte Urgull	ES	43.323155	-1.987114	NA	*12S+*	[19]
Bur2	Pl	Burgos	ES	42.35	-3.7	DQ081068	*12S+*	[13]
	Pl	Ulia	ES	43.326231	-1.953761	NA	*12S+*	[19]
Pht1	Pl	Tarragona	ES	41.116667	1.25	AF469438	*cybt+*	[8]
T1200	Pl	service area «El Montseny» along the Girona-Barcelona highway, 2 km NE of Llinar de Vallès	ES	41.6476	2.4249	KY461930	*ND4*	this study
Val1	Pl	Vall d Alinyá	ES	42.1828	1.42195	AF469442	*cybt+*	[8]
6.161	Pv	Albacete (ciudad)	ES	38.9782	-1.857717	HQ898112	*12S*	[11]
6.313	Pv	Riopar el Viejo	ES	38.499867	-2.418967	HQ898153	*12S*	[11]
6.317	Pv	Villanueva de Cordoba	ES	38.31535	-4.625317	HQ898160	*12S*	[11]
6.36	Pv	Virgen de la Cabeza, Andujar	ES	38.181833	-4.037983	HQ898162	*12S*	[11]
6.128	Pv	Cornalvo NP	ES	39.019867	-6.17615	HQ898125	*12S*	[11]
And10	Pv	Benatae	ES	38.35	-2.65	DQ081067	*12S+*	[13]
BEV.1900	Pv	Villanueva de los Escuderos	ES	40.0417	-2.3025	FJ208756	*D-loop*	[18]
BEV.1903	Pv	Road N.400 at Horcajada de la Torre (W. of Cuenca)	ES	40.032	-2.569	KY461892	*ND4*	this study
BEV.1910	Pv	Albalate de Zorita (N.NE. of Tarancón)	ES	40.308	-2.848	FJ208759	*D-loop*	[18]
BEV.1918	Pv	Gualda (N. of Sacedón)	ES	40.684	-2.686	FJ208762	*D-loop*	[18]
BEV.3992	Pv	"El Cotillo" estate, 12 km past Alcolea along road to embalse del Guadalmellato (16 km NE. of Córdoba)	ES	38.0103	-4.6499	FJ208726	*D-loop*	[18]
BEV.7371	Pv	Burunchel, in the village (Sierra de Cazorla)	ES	37.945	-2.954	FJ208727	*D-loop*	[18]
BEV.7373	Pv	Cañada de las Fuentes, Nacimiento del Rio Guadalquivir (Sierra de Cazorla)	ES	37.838	-2.972	KY461884	*D-loop*	this study
BEV.7403	Pv	Cazorla village, castle	ES	37.922	-2.99	FJ208721	*D-loop*	[18]
BEV.7525	Pv	service area 4 km SW. of Ciempozuelos (between Aranjuez and Valdemoro)	ES	40.134	-3.656	KY461929	*ND4*	this study

Sample code	Lineage[1]	Locality	Country[2]	Latitude	Longitude	GenBank accession number	Locus[3]	Reference[4]
CR1	Pv	Castanho del Robledo	ES	37.893667	-6.705933	EU269567	*ND4*	[15]
CV1	Pv	Castelo de Vide	PT	39.416667	-7.45	EU269562	*ND4*	[15]
DB11154	Pv	S of Arroyo Frio	ES	37.930861	-2.927944	KY461844	*16S*	this study
DB1728	Pv	Arroyo de la Luz	ES	39.483333	-6.583333	HQ898119	*12S*	[11]
DB1736	Pv	Cueva del Santillo	ES	37.924807	-2.952072	HQ898130	*12S*	[11]
DB1769	Pv	Cortijo de los Petrolos	ES	38.28569	-2.6485	HQ898127	*12S*	[11]
DB1776	Pv	Cortijo de Angelita	ES	38.312837	-2.619458	HQ898126	*12S*	[11]
DB1779	Pv	Area Recreativa de los Estrechos	ES	38.323485	-2.633985	HQ898117	*12S*	[11]
DB1781	Pv	Peña del Olivar	ES	38.371682	-2.578653	HQ898147	*12S*	[11]
DB1783	Pv	Cortijo El Maguillo	ES	38.240913	-2.779212	HQ898129	*12S*	[11]
DB1787	Pv	Fuente del Macho	ES	38.066492	-2.828838	HQ898134	*12S*	[11]
DB1800	Pv	Barranco del Buendesvío, Santa María del Espino	ES	40.972934	-2.25188	KY461842	*16S*	this study
DB1828	Pv	Piedra de los Endrinales	ES	38.533136	-2.413466	HQ898148	*12S*	[11]
DB1837	Pv	Fuente de Cueva Ahumada	ES	38.447088	-2.485636	HQ898133	*12S*	[11]
DB1862	Pv	Laguna de Arroyofrío	ES	38.411871	-2.512395	HQ898139	*12S*	[11]
DB1876	Pv	Palacio Gosalvez, Villalgordo del Júca	ES	39.296427	-2.070498	HQ898145	*12S*	[11]
DB1890	Pv	Río Linares- Riba de Saelices	ES	40.940402	-2.292685	HQ898152	*12S*	[11]
DB19833	Pv	Santa Maria da Feira castle	PT	40.920982	-8.543151	KP455498	*16S+*	[5]
DB2641	Pv	Mazarete	ES	40.925541	-2.165108	HQ898141	*12S*	[11]
DB2642	Pv	Cobeta	ES	40.884703	-2.137668	HQ898124	*12S*	[11]
DB2785	Pv	SW of Embalse del Tranco	ES	38.082325	-2.817273	HQ898157	*12S*	[11]
DB2862	Pv	Area Recreativa de Gil Cobo	ES	38.080451	-2.89991	HQ898114	*12S*	[11]
DB2866	Pv	Sierra de Segura, 5km W of Embalse del Tranco	ES	38.161933	-2.847769	HQ898156	*12S*	[11]
DB2911	Pv	Area Recreativa de Gil Cobo 2	ES	38.079387	-2.899079	HQ898116	*12S*	[11]
DB2959	Pv	E of Villanueva del Arzobispo	ES	38.180594	-2.914857	KY461843	*16S*	this study
DB2960	Pv	Rio Borosa (La Iruela)	ES	38.002384	-2.850674	HQ898149	*12S*	[11]
DB8904	Pv	Ourém	PT	39.691333	-8.583472	HQ898007	*ND4*	[11]
DB8905	Pv	Louriçal	PT	40.016806	-8.727389	HQ898006	*ND4*	[11]
DB9603	Pv	Olmeda de Cobeta	ES	40.866667	-2.183333	HQ898144	*12S*	[11]
DB9607	Pv	Monte Real	PT	39.85	-8.866667	HQ898142	*12S*	[11]
DB9647	Pv	Almoster	PT	39.241764	-8.794579	HQ898113	*12S*	[11]
DB9658	Pv	S. Pedro de Moel	PT	39.75	-9.016667	HQ898154	*12S*	[11]
DB9667	Pv	Casar de Cáceres	ES	39.566667	-6.416667	HQ898122	*12S*	[11]
DB9669	Pv	Castanheira de Pera	PT	40	-8.216667	HQ898123	*12S*	[11]
DB9676	Pv	El Laminador, Sierra de Aljubar	ES	38.483333	-2.4	HQ898132	*12S*	[11]
Ev4	Pv	Evora	PT	38.566667	-7.9	EU269564	*ND4*	[15]
Mad2	Pv	Madrid	ES	40.492967	-3.58675	AF469460	*cytb+*	[8]

Sample code	Lineage[1]	Locality	Country[2]	Latitude	Longitude	GenBank accession number	Locus[3]	Reference[4]
PH49	Pv	Rio Estena Hontanar	ES	39.561609	-4.589451	HQ898150	*12S*	[11]
PH50	Pv	Arroyo del Chorro - Los Navalucillos	ES	39.566383	-4.657822	HQ898120	*12S*	[11]
PH52	Pv	Rio Frio - Sevilleja de la Jara	ES	39.601025	-4.934205	HQ898151	*12S*	[11]
PH53	Pv	Fuente Vieja - Villarubia de Santiago	ES	39.989451	-3.346866	HQ898136	*12S*	[11]
PH55	Pv	Fuente Nueva - Villarubia de Santiago	ES	39.977923	-3.345485	HQ898135	*12S*	[11]
PH66	Pv	Ocaña	ES	39.963533	-3.504733	HQ898143	*12S*	[11]
PH72	Pv	Saelices	ES	39.919683	-2.804717	HQ898155	*12S*	[11]
PH76	Pv	Arroyo Brezoso	ES	39.360417	-4.358367	HQ898118	*12S*	[11]
PH80	Pv	Balazote	ES	38.8785	-2.2005	HQ898121	*12S*	[11]
PH87	Pv	El Chorro (Cabañeros NP)	ES	39.557983	-4.6537	HQ898131	*12S*	[11]
PH89	Pv	Fuertescusa	ES	40.475467	-2.177117	HQ898137	*12S*	[11]
PH91	Pv	Lagunas de la Ruidera	ES	39.061183	-3.006317	HQ898140	*12S*	[11]
PH95	Pv	La Roda	ES	39.200417	-2.165033	HQ898138	*12S*	[11]
PH98	Pv	Valencia del Ventoso	ES	38.2277	-6.483	HQ898159	*12S*	[11]
V1	Pv	Leiria	PT	39.746833	-8.809533	AF372063	*COI+*	[7]
V2	Pv	Portalegre	PT	39.283333	-7.433333	AF372064	*cytb*	[7]
V3	Pv	Beja	PT	38.016667	-7.866667	AF469455	*cytb+*	[7]
V4	Pv	Marvao	PT	39.4	-7.383333	AF372065	*COI*	[7]
V5	Pv	Agueda	PT	40.566667	-8.45	AF372066	*COI+*	[7]
SM2	Pv	S. Mamede	PT	39.467967	-7.634767	EU269563	*ND4*	[15]
Tie1	Pv	Tielmes	ES	40.233333	-3.316667	AY132323	*cytb+*	[10]
8.252	PvMA	Imlil	MA	31.100683	-7.91445	HQ898028	*ND4*	[11]
8.377	PvMA	Tislit Lake	MA	32.1964	-5.642933	HQ898032	*ND4*	[11]
AH1	PvMA	Ain Harhar	ALG	35.876417	1.940583	GQ856111	*12S+*	[12]
BEV.4450	PvMA	1 km past the Tizi-n-Tichka pass along the road to Ouarzazate at 2200 m a.s.l.	MA	31.2802	-7.3850	FJ764794	*D-loop*	[17]
Bt6	PvMA	Bab Taza	MA	35.022583	-5.204483	AY132324	*cytb+*	[10]
Cha22	PvMA	Charef	ALG	34.550467	2.796083	GQ856121	*12S+*	[12]
DB1024	PvMA	Agoudal	MA	32.03531	-5.46745	KY461868	*16S*	this study
DB1047	PvMA	Tizi-n-Tleta	MA	30.7809	-7.64354	HQ898229	*12S*	[11]
DB11145	PvMA	Y. Tichrout, Boulemane	MA	33.364548	-4.691156	KY461871	*16S*	this study
DB11153	PvMA	Zarka	MA	35.515723	-5.341238	KY461872	*16S*	this study
DB11705	PvMA	NW of Outerbat	MA	32.158112	-5.368369	KY461890	*ND4*	this study
DB1271	PvMA	Imouzzer Kandar to Annoceur	MA	33.62582	-4.896278	KY461870	*16S*	this study
DB1449	PvMA	Ceuta	ES*	35.888270	-5.316160	HQ898026	*ND4*	[11]
DB15554	PvMA	SE of Jnan Annich	MA	35.26684	-4.84243	KY461873	*16S*	this study
DB15571	PvMA	E of La Ville Nouvelle Cherafate power plant	MA	35.6632	-5.62317	KY461874	*16S*	this study

Sample code	Lineage[1]	Locality	Country[2]	Latitude	Longitude	GenBank accession number	Locus[3]	Reference[4]
DB1611	PvMA	road to Jbel Siroua	MA	30.77671	-7.652988	HQ898228	*12S*	[11]
DB3878	PvMA	S of Ain Leuh	MA	33.1863	-5.33132	KY461867	*16S*	this study
DB442	PvMA	S of Tifni	MA	31.5892	-6.936883	KY461866	*16S*	this study
DB5040	PvMA	Minas (Tadert/ Tizin Tichka) on the way down	MA	31.30059	-7.39149	HQ898030	*ND4*	[11]
DB5042	PvMA	Jbel Owlime	MA	30.81708	-8.862983	HQ898029	*ND4*	[11]
DB5048	PvMA	base of the mountains (Minas)	MA	31.29504	-7.376495	HQ898024	*ND4*	[11]
DB5087	PvMA	Talzemt	MA	33.64037	-4.2007	HQ898031	*ND4*	[11]
DB6829	PvMA	E of Anergui	MA	32.062109	-5.888981	KY461889	*ND4*	this study
DB86	PvMA	E of Sidi Yahya	MA	32.6708	-5.453117	KY461865	*16S*	this study
DB8719	PvMA	Debdou	MA	33.872467	-3.038783	EF081087	*ND4*	[14]
DB8742	PvMA	Ketama	MA	34.878233	-4.610867	EF081105	*ND4*	[14]
DB8755	PvMA	Jbel Tazzeka	MA	34.1042	-4.0725	EF081085	*ND4*	[14]
DB8843	PvMA	Ifrane	MA	33.533332	-5.116664	HQ898027	*ND4*	[11]
DB8876	PvMA	Tizi nTest	MA	30.833333	-8.333333	EF081102	*ND4*	[14]
DB8902	PvMA	Balcon d Ito	MA	33.516667	-5.283333	EF081098	*ND4*	[14]
DB8906	PvMA	Chefchaouen	MA	35.166667	-5.266667	EF081106	*ND4*	[14]
DB8911	PvMA	Foum Kheneg	MA	33.155217	-5.068267	EF081090	*ND4*	[14]
DB9130	PvMA	Beni Amint	MA	33.6468	-4.08852	HQ898025	*ND4*	[11]
DB972	PvMA	Anefgou	MA	32.302662	-5.327832	KY461869	*16S*	this study
Djek17	PvMA	Djebel Ksel	ALG	33.733233	1.1691	GQ856123	*12S+*	[12]
Dju939	PvMA	Djurjura	ALG	36.47105	3.996633	GQ856115	*12S+*	[12]
E16081	PvMA	N Oukaimeden	MA	31.20355	-7.86172	AY132326	*cytb+*	[10]
E29051	PvMA	8Km SW Zinat	MA	35.383333	-5.466667	AY132329	*cytb+*	[10]
E290510	PvMA	El-Had	MA	35	-5.383333	AY132337	*cytb+*	[10]
E29053	PvMA	15 Km SW Zinat	MA	35.316667	-5.483333	AY132331	*cytb+*	[10]
E29056	PvMA	Bab-Berred	MA	35.006533	-4.898633	AY132334	*cytb+*	[10]
E29058	PvMA	Taza	MA	34.22119	-4.015857	AY132335	*cytb+*	[10]
E31051	PvMA	Jbel Musa	MA	35.866667	-5.4	AY132344	*cytb+*	[10]
E31052	PvMA	Azrou	MA	33.416667	-5.2	AY132345	*cytb+*	[10]
Mch1	PvMA	M Chedallarh	ALG	36.3688	4.2708	GQ856113	*12S+*	[12]
MisD	PvMA	Mischliffen	MA	33.405433	-5.103317	AY132318	*cytb+*	[10]
MVZ 178291	PvMA	El Ksiba	MA	32.558531	-6.070278	AY234156	*cytb+*	[2]
MVZ 186233	PvMA	Ksar Es-Srhir	ES*	35.849624	-5.563936	AY234153	*cytb+*	[2]
Ouk7	PvMA	Oukaïmeden	MA	31.20355	-7.861717	AY132321	*cytb+*	[10]
Tah1	PvMA	Tahament	ALG	36.38326	5.0581	GQ856109	*12S+*	[12]
Tia1	PvMA	Tiaret	ALG	35.293066	1.263117	GQ856117	*12S+*	[12]
Tlem15	PvMA	Tlemcen	ALG	34.84285	-1.292167	GQ856119	*12S+*	[12]
	PvMA	10Km W Bab-Berred	MA	35.023881	-4.983403	AJ250167	*12S*	[16]
	PvMA	Midelt	MA	32.682367	-4.74265	EF081088	*ND4*	[14]
8.122	PvSS	Peñarroya-Pueblonuevo	ES	38.29805	-5.264933	HQ898020	*ND4*	[11]

Sample code	Lineage[1]	Locality	Country[2]	Latitude	Longitude	GenBank accession number	Locus[3]	Reference[4]
8.26	PvSS	Matalascañas	ES	37.253617	-6.561182	HQ898019	*ND4*	[11]
8.57	PvSS	Torcal de Antequera	ES	36.954717	-4.544417	HQ898023	*ND4*	[11]
8.69	PvSS	Playa de la Vibora	ES	36.49005	-4.775733	HQ898021	*ND4*	[11]
8.119	PvSS	Córdoba city	ES	37.87375	-4.7865	HQ898016	*ND4*	[11]
Ate1	PvSS	Barbate	ES	36.189717	-5.921667	AY132316	*cytb+*	[10]
BEV.3902	PvSS	Sierra Nevada, road towards Pico Veleta at 1300 m a.s.l.	ES	37.1422	-3.4871	KY461896	*ND4*	[17]
BEV.3908	PvSS	Sierra Nevada, road towards Pico Veleta around 1800-2000 m a.s.l.	ES	37.1094	-3.4293	KY461897	*ND4*	this study
BEV.4588	PvSS	El Rocio (Coto Doñana)	ES	37.136	-6.475	FJ764795	*D-loop*	[17]
BEV.4598	PvSS	crossing of roads A.369 and MA.3707 (S of Ronda)	ES	36.675	-5.203	KY461883	*D-loop*	this study
BEV.4603	PvSS	along road A.366, 5 km past the crossing with road A.376 going towards north (NE of Ronda)	ES	36.757	-5.111	FJ764798	*D-loop*	[17]
BEV.8155	PvSS	6 km past Pinos del Valle on road to Guájar-Faragüit (NNW. Motril)	ES	36.8822	-3.5395	FJ764801	*D-loop*	[17]
Cin1	PvSS	Guadalcacin	ES	36.633333	-5.65	AY132322	*cytb+*	[10]
DB11134	PvSS	Rota	ES	36.619303	-6.3635	KY461876	*16S*	this study
DB11142	PvSS	Barriada de las Flores, Jerez de la Frontera	ES	36.717013	-6.100507	KY461875	*16S*	this study
DB1251	PvSS	Castro del Rio	ES	37.683198	-4.474988	HQ898014	*ND4*	[11]
DB1380	PvSS	Granada city	ES	37.176377	-3.592951	HQ898017	*ND4*	[11]
DB1390	PvSS	Sierra de Aljibe	ES	36.97299	-5.663125	HQ898022	*ND4*	[11]
DB1811	PvSS	Bonanza	ES	36.800000	-6.333333	HQ898223	*12S*	[11]
DB1874	PvSS	Linares	ES	38.09362	-3.635844	HQ898225	*12S*	[11]
DB3871	PvSS	Sanlucar La Mayor	ES	37.386906	-6.20347	HQ898227	*12S*	[11]
DB446	PvSS	Parque Nacional Los Alcornocales	ES	36.35	-5.533333	HQ898226	*12S*	[11]
DB51	PvSS	(southern Cadiz) aka “Castillo de la Duquesa”	ES	36.351679	-5.233470	HQ898224	*12S*	[11]
DB8786	PvSS	La Barrosa	ES	36.3733	-6.18695	EF081076	*ND4*	[14]
E16084	PvSS	Mairena del Aljarafe	ES	37.316667	-6.066667	AY132327	*cytb+*	[10]
E16085	PvSS	Sevilla	ES	37.377217	-5.98695	AY132328	*cytb+*	[10]
Elv1	PvSS	Huelva	ES	37.25	-6.95	AY132317	*cytb+*	[10]
MNCN 11088	PvSS	Rio Hozgarganta	ES	36.443417	-5.451233	AY234162	*cytb+*	[2]
	PvSS	Facinas	ES	36.133333	-5.7	AY234159	*cytb+*	[2]

[1]Lineage names correspond to: Pb, *P. bocagei*; Pc, *P. carbonelli*; Pgl, *P. guadarramae lusitanicus*; Pgg, *P. g. guadarrramae; Pv, P. virescens*; Pl, *P. liolepis*; PhG, *P. hispanicus* Galera; V, Valencia lineage; PhAM, *P. hispanicus* Albacete/Murcia; PvSS, *P. vaucheri* Southern Spain; PvMA, *P. vaucheri* Morocco/Algeria; JS, Jbel Siroua lineage; TAg, Tunisia/Algeria group (including TAg[a], PHTA; TAg[b], PHAZA; TAg[c], PHBAT from [11]). [2]Contries: PT, Portugal; ES, Spain

(ES*, North of Africa); FR, France; AD, Andorra; MA, Morocco; ALG, Algeria, TN, Tunisia. [3]*12S*: *12S ribosomal RNA* gene; *16S*: *16S ribosomal* RNA gene; *cytb*: *cytochrome b* gene fragment; *ND4*: *NADH dehydrogenase subunit 4* gene fragment (and adjacent tRNAs); *D-loop*: control region gene fragment; *COI*: *cytochrome c oxidase subunit I* gene fragment. A "+" indicates that also other mitochondrial fragments were amplified for each sample (see corresponding reference). [4]References: [1]Arntzen and Sá-Sousa P 2007; [2]Busack et al. 2005; [3]Carranza et al. 2004; [4]Castilla et al. 1998; [5]Dias et al. 2016; [6]Harris and Arnold 1999; [7]Harris and Sá-Sousa 2001; [8]Harris and Sá-Sousa 2002; [9]Harris et al. 2002a; [10]Harris et al. 2002b; [11]Kaliontzopoulou et al. 2011; [12]Lima et al. 2009; [13]Pinho et al. 2006; [14]Pinho et al. 2007; [15]Pinho et al.2008; [16]Oliverio et al. 2000; [17]Renoult 2006; [18]Renoult et al. 2009; [19]Sanz-Azkue et al. 2006.

Table S1.2. Pairwise Jaccard's similarity index (JSI) for geographic distribution overlap between each *Podarcis hispanicus* complex form.

	Pb	**Pc**	**Pgl**	**Pgg**	**Pv**	**Pl**	**PhG**	**PhAM**	**V**	**PvSS**	**PvMA**	**JS**
Pc	0.053											
Pgl	0.483	0.074										
Pgg	0	0.038	0.052									
Pv	0.026	0.028	0.043	0.117								
Pl	0.026	0	0.057	0.009	0.049							
PhG	0	0	0	0	0.063	0						
PhAM	0	0	0	0	0.059	0	0					
V	0	0	0	0	0.126	0.082	0.339	0.182				
PvSS	0	0.014	0	0	0.097	0	0.056	0	0.048			
PvMA	0	0	0	0	0	0	0	0	0	0		
JS	0	0	0	0	0	0	0	0	0	0	0.107	
PhTA	0	0	0	0	0	0	0	0	0	0	0.023	0

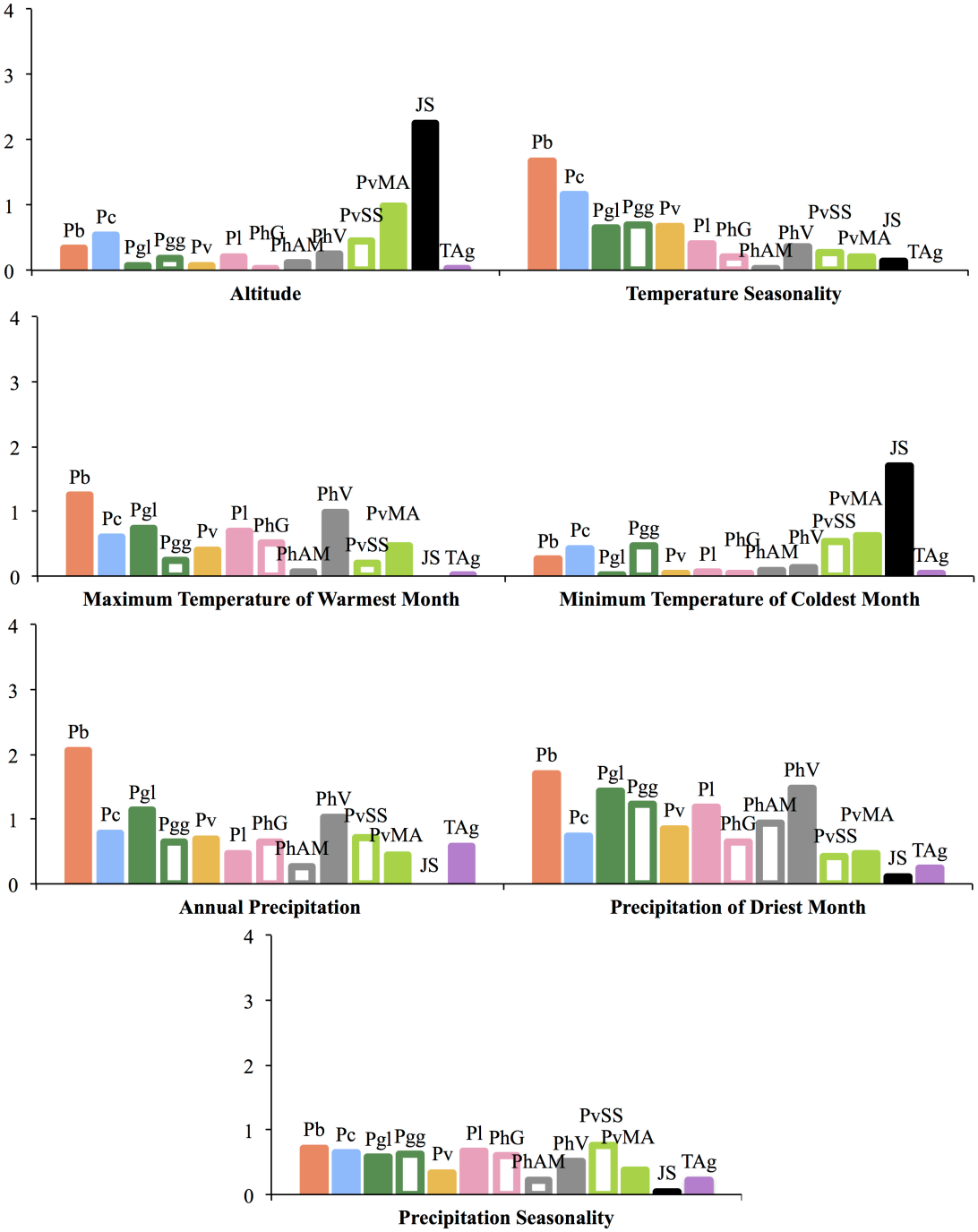

Figure S1.1. Results of the jack-knife tests for Maxent models gain with only each variable. Colours and species acronyms as in Figure 1.1. and Table S1.1. above.

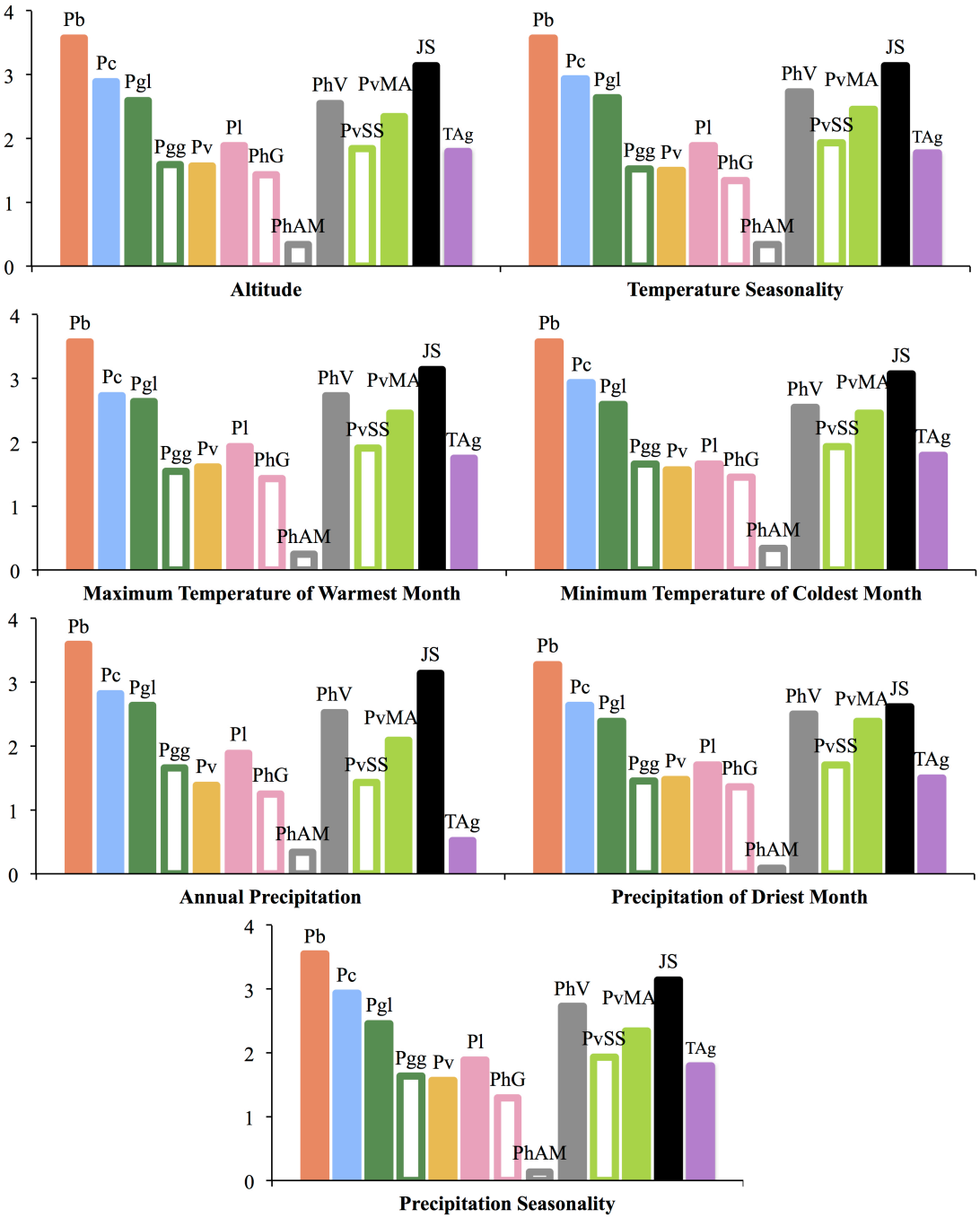

Figure S1.2. Results of the jackknife tests for Maxent models gain excluding only one variable. Colours and species acronyms as in Figure 1.1. and Table S1.1. above.

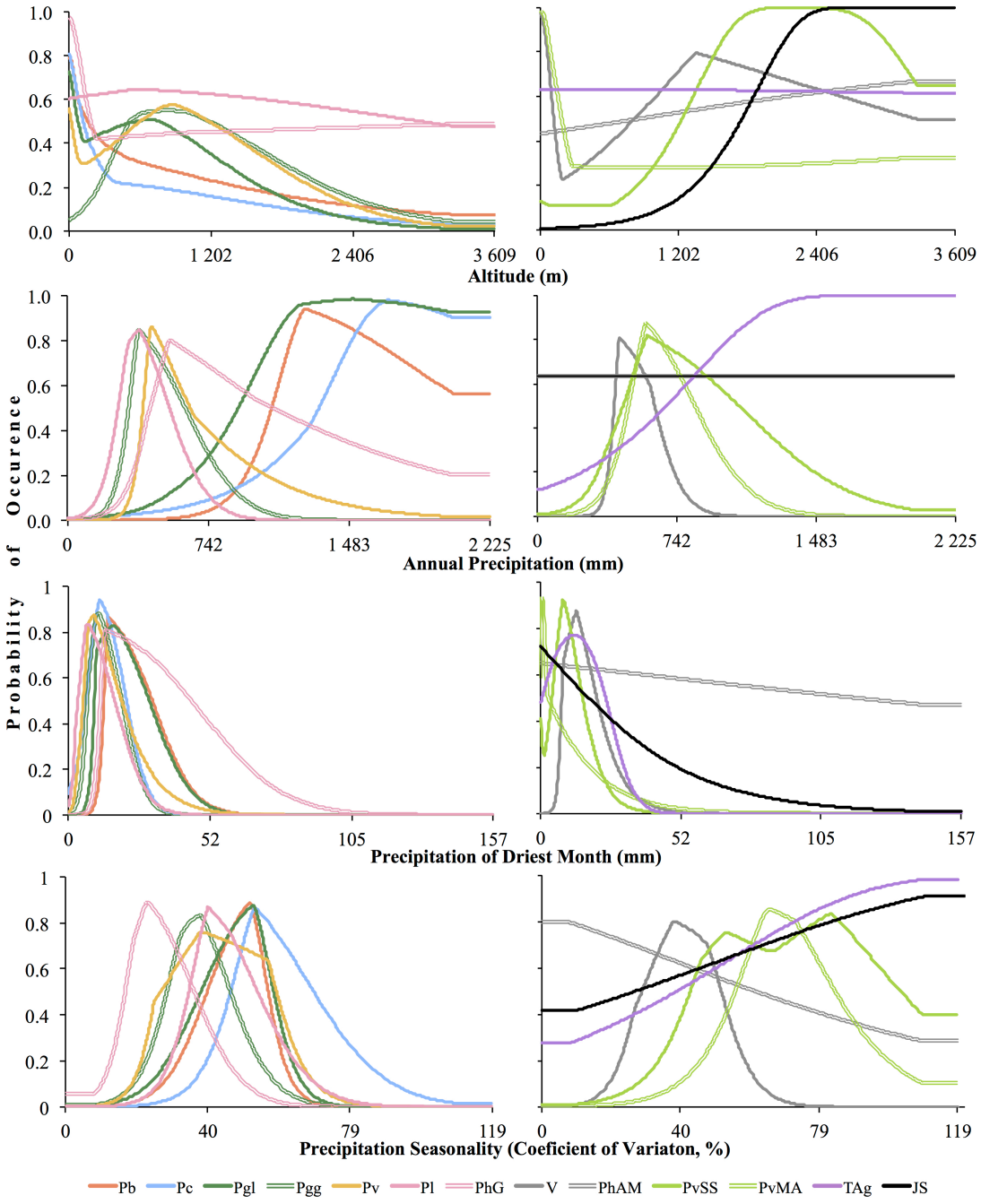

Figure S1.3. Response curves of variables used to calculate the average Maxent models for *Podarcis* lizards. Each plot represents variation of logistic prediction values across the observed range in the study area for each environmental variable, while keeping all other environmental variables at their average sample value. Colours and species acronyms as in Figure 1.1. and Table S1.1. above.

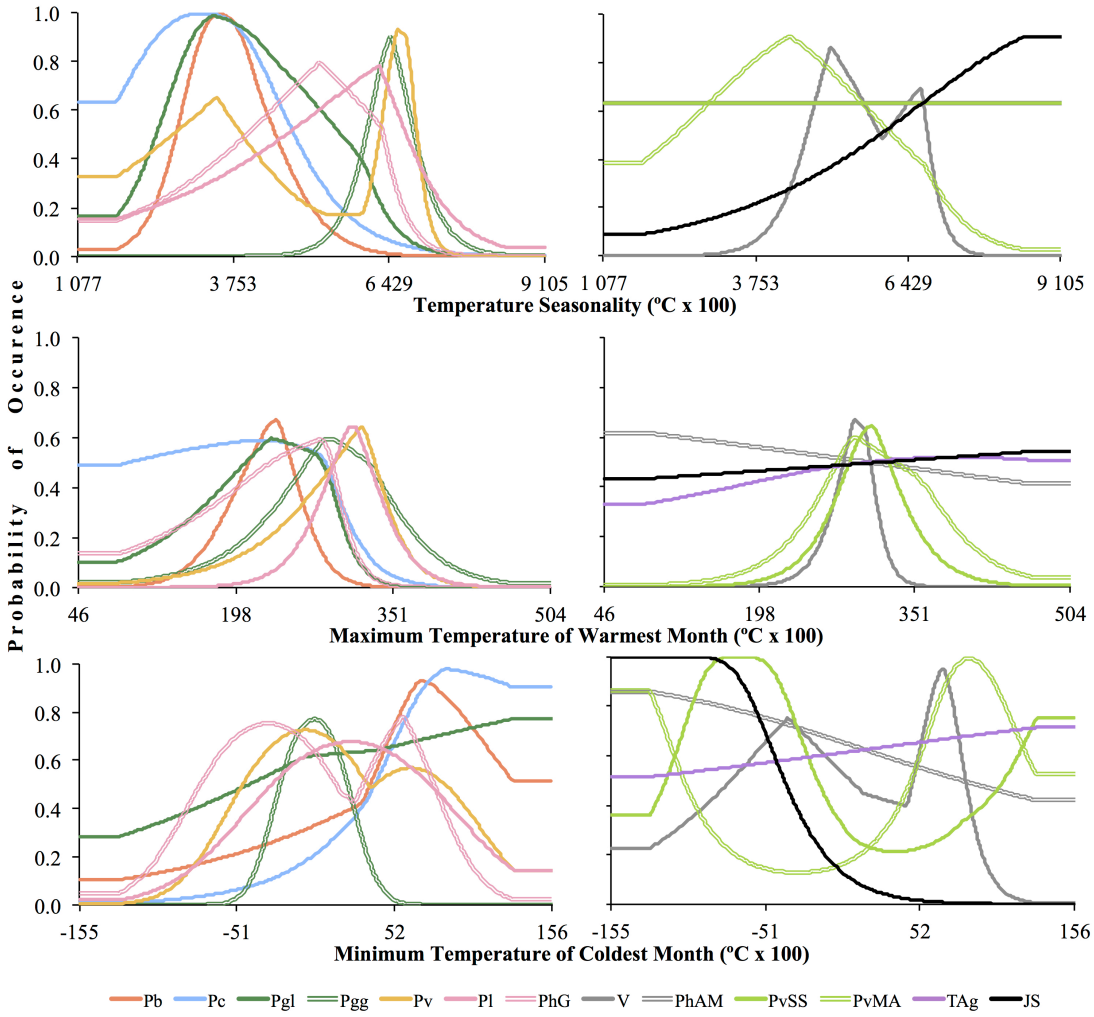

Figure S1.3. (Cont.)

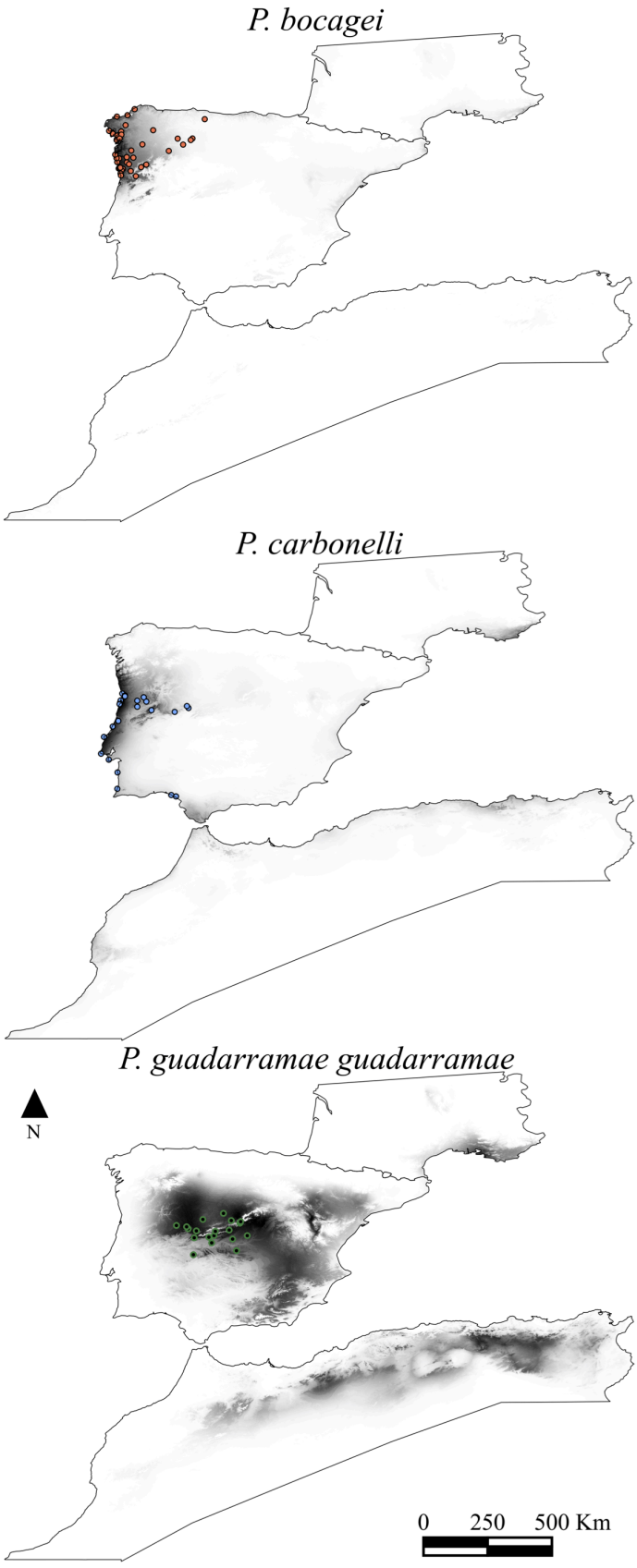

Figure S1.4. Maxent models for each form of the *Podarcis hispanicus* complex with the presence records used for Maxent modelling. Dark and light colours represent areas of high and low suitability, respectively.

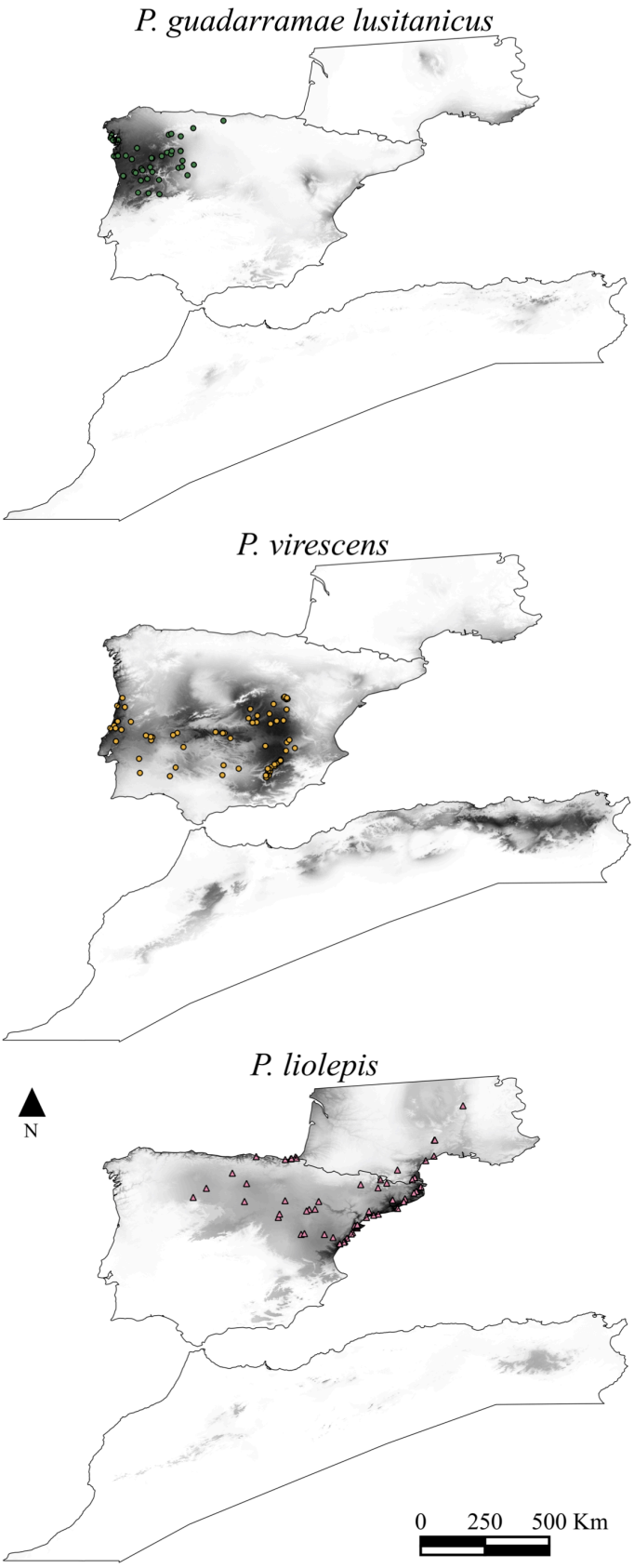

Figure S1.4. (Cont.)

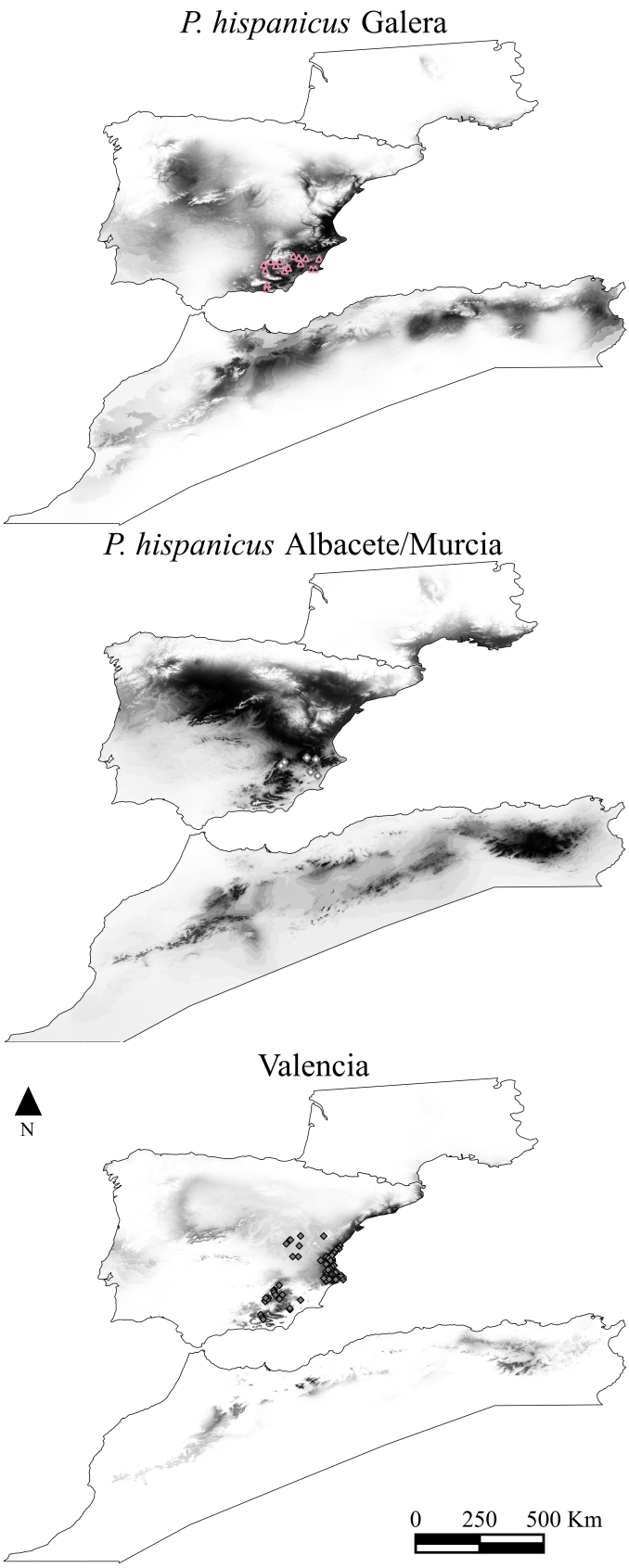

Figure S1.4. (Cont.)

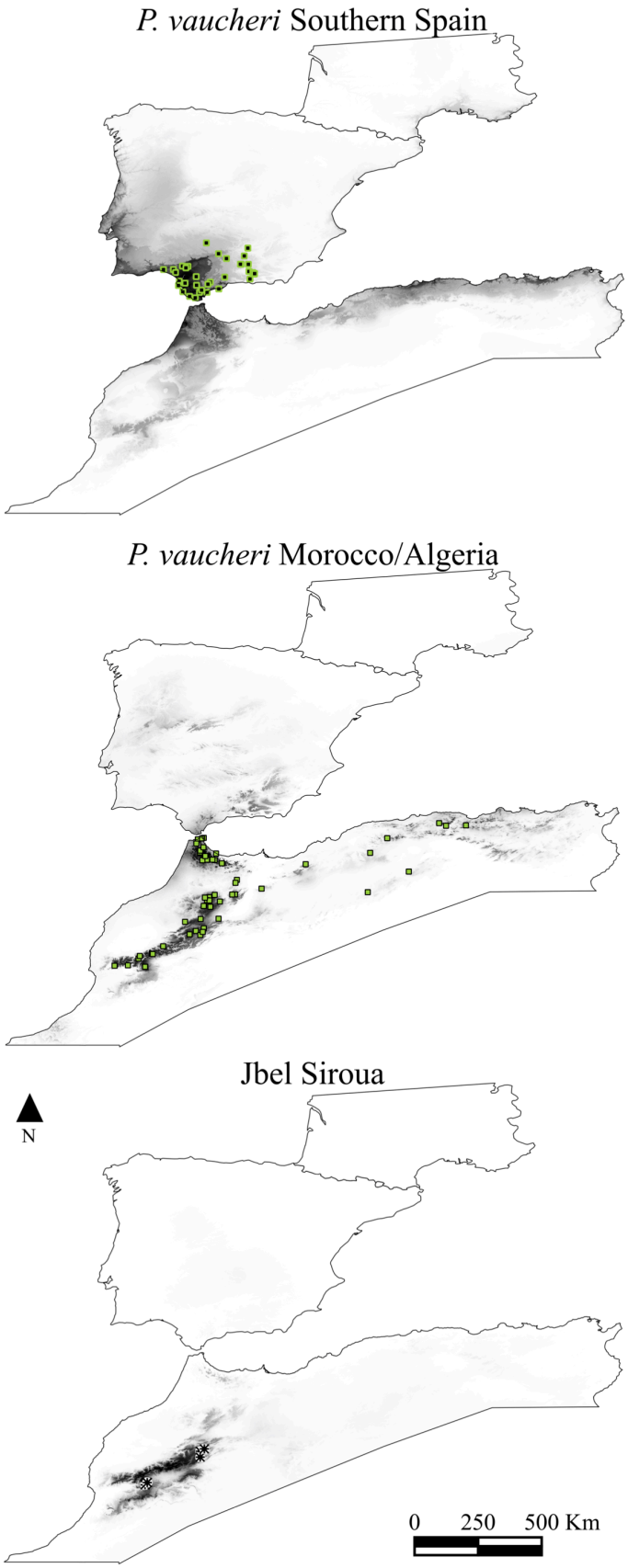

Figure S1.4. (Cont.)

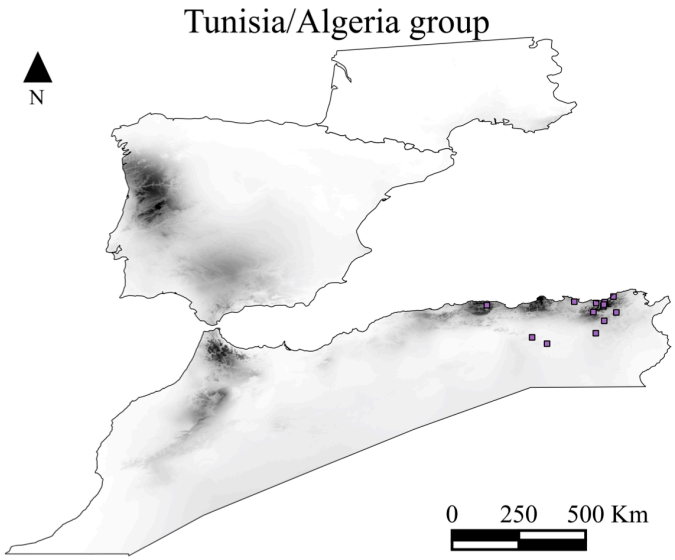

Figure S1.4. (Cont.)

Appendix II. Genome-wide patterns of interspecific admixture in a natural hybrid zone in late stages of speciation – Supporting Information

Table S2.1. Fisher's exact test *p*-values for comparisons among the average of each simulated scenario (SS) and observed data (OD) in each scenario and the number of similar simulations to OD in each SS. OD are the classes of 10% of distribution of admixture proportions in the centre of the hybrid zone and the average SS is the average between the 100 simulations for each combination of random (RM) and non-random mating (NRM) with several immigration rates (0, 0.2, 0.3 and 0.4). * indicate the non-significant tests after Bonferroni correction (significant at *p*-value < 6.25×10^{-3}). Number of simulations (out of 100) that are not significantly different from OD in each SS (tests were considered significant at *p*-value < 5.0×10^{-4} after Bonferroni correction).

Comparison	*p*-value	Nr of simulations in each SS that do not significantly differ from OD
OD : RM0	1.50×10^{-64}	0
OD : RM0.2	9.98×10^{-13}	0
OD : RM0.3	1.87×10^{-07}	1
OD : RM0.4	0.000171	34
OD : NMR0	6.08×10^{-64}	0
OD : NMR0.2	2.54×10^{-11}	0
OD : NMR0.3	1.91×10^{-05}	11
OD : NMR0.4	0.052*	91

Table S2.2. Significance of the proportion of ratios F1/BC and Pure/Total similar to observed data for each simulated scenario. Significant differences are marked with *.

Ratio	Simulation	*p*-value
F1/BC	RM0	0
	RM0.05	0
	RM0.2	0
	RM0.3	0.11*
	RM0.4	0.37*
	NRM0	0
	NRM0.05	0
	NRM0.2	0
	NRM0.3	0.04
	NRM0.4	0.15*
Pure/Total	RM0	0
	RM0.05	0
	RM0.2	0
	RM0.3	0
	RM0.4	0
	NRM0	0
	NRM0.05	0
	NRM0.2	0
	NRM0.3	0
	NRM0.4	0

Table S2.3. Number of loci fitted in each tested model with HZAR.

Model	Nr of loci
none_none	497
fixed_none	174
free_none	687
none_mirror	161
fixed_mirror	44
free_mirror	4
none_right	508
fixed_right	46
free_right	82
none_left	61
fixed_left	6
free_left	2
none_both	10
fixed_both	0
free_both	0

Table S2.4. Fisher's exact tests *p*-values on each of the seven comparisons between the distribution of loci with significant parameters in the Z chromosome and in the autosomes. Significant differences are marked with *.

Significant set of loci	Proportion of loci in the Z chromosome	Proportion of loci in autosomes	*p*-values
Alpha negative	0.41	0.31	7.72 x 10-3*
Alpha positive	0.11	0.22	4.87 x 10-6*
Beta negative	0.14	0.22	2.21 x 10-3*
Beta positive	0.14	0.18	0.048*
c shifted north	0.01	0.006	0.312
c shifted south	0.005	0.006	1
w higher	0.02	0.03	0.624

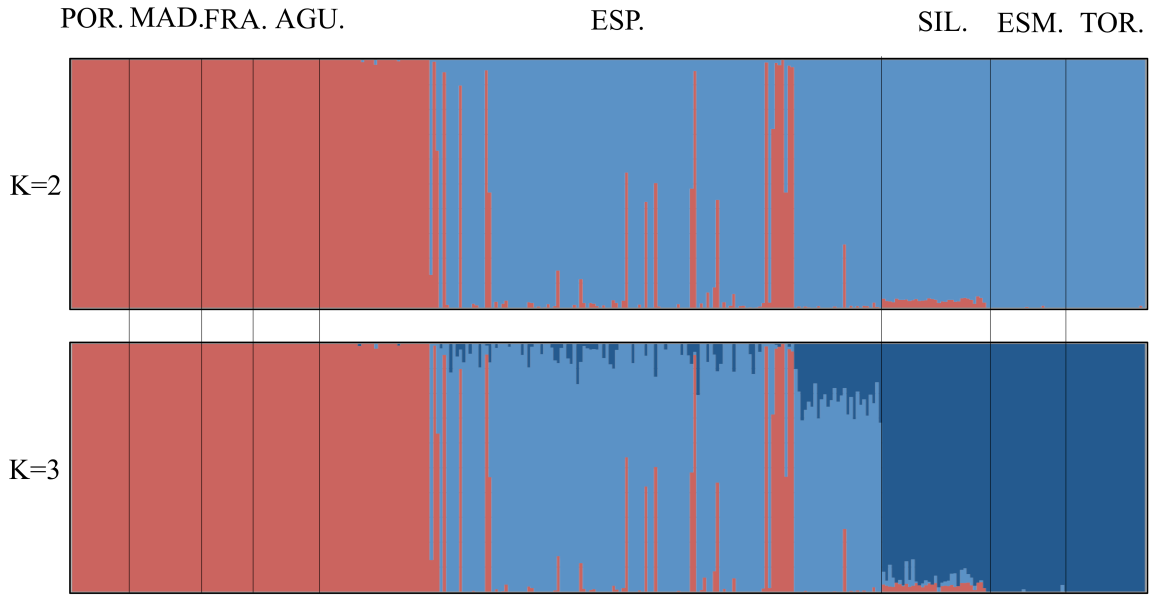

Figure S2.1. Results from individual multilocus genotype clustering computed with Admixture for the 6905 loci. Each individual is represented by a vertical line proportionally partitioned into the K=2 (above) and K=3 (best K value, below) coloured segments. Population acronyms and colors as in Figure 3.1.

Appendix III. Evolution of sympatry without complete reproductive isolation: does hybridization matter for *Podarcis carbonelli* conservation? – Supporting Information

Table S3.1. Samples used to analyze each *Podarcis* contact zone, with corresponding morphological identification, geographical coordinates and the place where the samples are stored. The contact zones analyzed are identified as 1 – *P. bocagei* vs *P. carbonelli*, 2 – *P. virescens* vs *P. carbonelli*, 3 – *P. guadarramae lusitanicus* vs *P. carbonelli* and 4 – *P. vaucheri* SSp vs *P. carbonelli*; CZ are the samples from the contact zone and REF are the samples used as references.

Sample label	Contact zone analyzed[1]	Field ID	Latitude	Longitude	Stored at
DB19668	1 (CZ)	*P. bocagei*	41.031	-8.646	CIBIO-InBio
DB19607	1 (CZ)	*P. bocagei*	41.031	-8.645	CIBIO-InBio
DB19682	1 (CZ)	*P. bocagei*	41.031	-8.645	CIBIO-InBio
DB19681	1 (CZ)	*P. carbonelli*	41.030	-8.645	CIBIO-InBio
DB19683	1 (CZ)	*P. carbonelli*	41.030	-8.645	CIBIO-InBio
DB19552	1 (CZ)	*P. carbonelli*	41.030	-8.646	CIBIO-InBio
DB19555	1 (CZ)	*P. carbonelli*	41.030	-8.645	CIBIO-InBio
DB21509	1 (CZ)	*P. sp.*	41.029	-8.645	CIBIO-InBio
DB19602	1 (CZ)	*P. carbonelli*	41.029	-8.645	CIBIO-InBio
DB21508	1 (CZ)	*P. carbonelli*	41.029	-8.645	CIBIO-InBio
DB19579	1 (CZ)	*P. carbonelli*	41.029	-8.645	CIBIO-InBio
DB19598	1 (CZ)	*P. carbonelli*	41.029	-8.645	CIBIO-InBio
DB19600	1 (CZ)	*P. carbonelli*	41.029	-8.645	CIBIO-InBio
DB19588	1 (CZ)	*P. carbonelli*	41.029	-8.645	CIBIO-InBio
DB19551	1 (CZ)	*P. carbonelli*	41.029	-8.645	CIBIO-InBio
DB19563	1 (CZ)	*P. carbonelli*	41.029	-8.645	CIBIO-InBio
DB19553	1 (CZ)	*P. carbonelli*	41.029	-8.645	CIBIO-InBio
DB21505	1 (CZ)	*P. sp.*	41.029	-8.645	CIBIO-InBio
DB19582	1 (CZ)	*P. bocagei*	41.029	-8.645	CIBIO-InBio
DB19585	1 (CZ)	*P. carbonelli*	41.029	-8.646	CIBIO-InBio
DB19566	1 (CZ)	*P. carbonelli*	41.029	-8.645	CIBIO-InBio
DB19556	1 (CZ)	*P. carbonelli*	41.029	-8.646	CIBIO-InBio
DB19945	1 (CZ)	*P. bocagei*	41.029	-8.645	CIBIO-InBio
DB21513	1 (CZ)	*P. sp.*	41.029	-8.645	CIBIO-InBio
DB21511	1 (CZ)	*P. sp.*	41.029	-8.645	CIBIO-InBio
DB19559	1 (CZ)	*P. carbonelli*	41.029	-8.645	CIBIO-InBio
DB21512	1 (CZ)	*P. sp.*	41.029	-8.645	CIBIO-InBio
DB19564	1 (CZ)	*P. carbonelli*	41.029	-8.646	CIBIO-InBio
DB19554	1 (CZ)	*P. carbonelli*	41.029	-8.645	CIBIO-InBio
DB19929	1 (CZ)	*P. sp.*	41.029	-8.645	CIBIO-InBio
DB21510	1 (CZ)	*P. sp.*	41.029	-8.645	CIBIO-InBio
DB19712	1 (CZ)	*P. carbonelli*	41.028	-8.645	CIBIO-InBio
DB21503	1 (CZ)	*P. sp.*	41.028	-8.645	CIBIO-InBio
DB19949	1 (CZ)	*P. sp.*	41.028	-8.645	CIBIO-InBio

Sample label	Contact zone analyzed[1]	Field ID	Latitude	Longitude	Stored at
DB19720	1 (CZ)	*P. carbonelli*	41.028	-8.646	CIBIO-InBio
DB19597	1 (CZ)	*P. carbonelli*	41.028	-8.645	CIBIO-InBio
DB19584	1 (CZ)	*P. carbonelli*	41.028	-8.645	CIBIO-InBio
DB19947	1 (CZ)	*P. sp.*	41.028	-8.645	CIBIO-InBio
DB19924	1 (CZ)	*P. sp.*	41.028	-8.646	CIBIO-InBio
DB19587	1 (CZ)	*P. carbonelli*	41.028	-8.645	CIBIO-InBio
DB19946	1 (CZ)	*P. sp.*	41.028	-8.645	CIBIO-InBio
DB19557	1 (CZ)	*P. carbonelli*	41.028	-8.645	CIBIO-InBio
DB21502	1 (CZ)	*P. carbonelli*	41.028	-8.645	CIBIO-InBio
DB19711	1 (CZ)	*P. carbonelli*	41.028	-8.645	CIBIO-InBio
DB19569	1 (CZ)	*P. carbonelli*	41.028	-8.646	CIBIO-InBio
DB19580	1 (CZ)	*P. carbonelli*	41.028	-8.646	CIBIO-InBio
DB19935	1 (CZ)	*P. sp.*	41.028	-8.646	CIBIO-InBio
DB19593	1 (CZ)	*P. carbonelli*	41.028	-8.646	CIBIO-InBio
DB19562	1 (CZ)	*P. carbonelli*	41.028	-8.646	CIBIO-InBio
DB19936	1 (CZ)	*P. sp.*	41.028	-8.645	CIBIO-InBio
DB19586	1 (CZ)	*P. carbonelli*	41.028	-8.645	CIBIO-InBio
DB19595	1 (CZ)	*P. carbonelli*	41.028	-8.645	CIBIO-InBio
DB19940	1 (CZ)	*P. carbonelli*	41.028	-8.644	CIBIO-InBio
DB19942	1 (CZ)	*P. carbonelli*	41.028	-8.644	CIBIO-InBio
DB21514	1 (CZ)	*P. sp.*	41.028	-8.645	CIBIO-InBio
DB19567	1 (CZ)	*P. carbonelli*	41.028	-8.645	CIBIO-InBio
DB19561	1 (CZ)	*P. carbonelli*	41.028	-8.645	CIBIO-InBio
DB19571	1 (CZ)	*P. carbonelli*	41.028	-8.645	CIBIO-InBio
DB19925	1 (CZ)	*P. sp.*	41.028	-8.645	CIBIO-InBio
DB19944	1 (CZ)	*P. carbonelli*	41.028	-8.645	CIBIO-InBio
DB19592	1 (CZ)	*P. carbonelli*	41.028	-8.645	CIBIO-InBio
DB19572	1 (CZ)	*P. carbonelli*	41.028	-8.646	CIBIO-InBio
DB19926	1 (CZ)	*P. sp.*	41.027	-8.645	CIBIO-InBio
DB19581	1 (CZ)	*P. carbonelli*	41.027	-8.655	CIBIO-InBio
DB21501	1 (CZ)	*P. carbonelli*	41.027	-8.645	CIBIO-InBio
DB19699	1 (CZ)	*P. carbonelli*	41.027	-8.646	CIBIO-InBio
DB19932	1 (CZ)	*P. sp.*	41.027	-8.645	CIBIO-InBio
DB19698	1 (CZ)	*P. carbonelli*	41.027	-8.646	CIBIO-InBio
DB19591	1 (CZ)	*P. carbonelli*	41.027	-8.645	CIBIO-InBio
DB19599	1 (CZ)	*P. carbonelli*	41.027	-8.645	CIBIO-InBio
DB19719	1 (CZ)	*P. bocagei*	41.027	-8.646	CIBIO-InBio

Sample label	Contact zone analyzed[1]	Field ID	Latitude	Longitude	Stored at
DB19603	1 (CZ)	*P. carbonelli*	41.027	-8.645	CIBIO-InBio
DB19565	1 (CZ)	*P. carbonelli*	41.027	-8.645	CIBIO-InBio
DB19590	1 (CZ)	*P. carbonelli*	41.027	-8.645	CIBIO-InBio
DB19594	1 (CZ)	*P. carbonelli*	41.027	-8.645	CIBIO-InBio
DB19938	1 (CZ)	*P. sp.*	41.027	-8.645	CIBIO-InBio
DB19927	1 (CZ)	*P. sp.*	41.027	-8.644	CIBIO-InBio
DB19576	1 (CZ)	*P. carbonelli*	41.027	-8.645	CIBIO-InBio
DB21507	1 (CZ)	*P. carbonelli*	41.027	-8.645	CIBIO-InBio
DB19589	1 (CZ)	*P. carbonelli*	41.027	-8.645	CIBIO-InBio
DB19568	1 (CZ)	*P. carbonelli*	41.027	-8.645	CIBIO-InBio
DB19558	1 (CZ)	*P. carbonelli*	41.027	-8.645	CIBIO-InBio
DB19708	1 (CZ)	*P. bocagei*	41.027	-8.643	CIBIO-InBio
DB19709	1 (CZ)	*P. bocagei*	41.027	-8.643	CIBIO-InBio
DB19596	1 (CZ)	*P. carbonelli*	41.027	-8.644	CIBIO-InBio
DB19715	1 (CZ)	*P. carbonelli*	41.027	-8.645	CIBIO-InBio
DB19716	1 (CZ)	*P. carbonelli*	41.027	-8.645	CIBIO-InBio
DB19928	1 (CZ)	*P. sp.*	41.027	-8.645	CIBIO-InBio
DB19560	1 (CZ)	*P. carbonelli*	41.027	-8.645	CIBIO-InBio
DB19583	1 (CZ)	*P. carbonelli*	41.027	-8.645	CIBIO-InBio
DB19601	1 (CZ)	*P. carbonelli*	41.027	-8.645	CIBIO-InBio
DB21504	1 (CZ)	*P. sp.*	41.027	-8.644	CIBIO-InBio
DB19934	1 (CZ)	*P. sp.*	41.027	-8.645	CIBIO-InBio
DB19697	1 (CZ)	*P. carbonelli*	41.027	-8.645	CIBIO-InBio
DB19577	1 (CZ)	*P. carbonelli*	41.027	-8.645	CIBIO-InBio
DB19948	1 (CZ)	*P. sp.*	41.027	-8.645	CIBIO-InBio
DB19578	1 (CZ)	*P. carbonelli*	41.027	-8.645	CIBIO-InBio
DB21500	1 (CZ)	*P. carbonelli*	41.027	-8.644	CIBIO-InBio
DB19930	1 (CZ)	*P. sp.*	41.027	-8.645	CIBIO-InBio
DB19574	1 (CZ)	*P. carbonelli*	41.026	-8.645	CIBIO-InBio
DB19931	1 (CZ)	*P. sp.*	41.026	-8.645	CIBIO-InBio
DB19933	1 (CZ)	*P. sp.*	41.026	-8.644	CIBIO-InBio
DB19723	1 (CZ)	*P. carbonelli*	41.026	-8.645	CIBIO-InBio
DB19701	1 (CZ)	*P. carbonelli*	41.026	-8.645	CIBIO-InBio
DB19727	1 (CZ)	*P. bocagei*	41.026	-8.645	CIBIO-InBio
DB19705	1 (CZ)	*P. carbonelli*	41.026	-8.645	CIBIO-InBio
DB19726	1 (CZ)	*P. bocagei*	41.026	-8.644	CIBIO-InBio
DB19707	1 (CZ)	*P. bocagei*	41.026	-8.644	CIBIO-InBio
DB19724	1 (CZ)	*P. bocagei*	41.026	-8.644	CIBIO-InBio

Sample label	Contact zone analyzed[1]	Field ID	Latitude	Longitude	Stored at
DB19718	1 (CZ)	*P. bocagei*	41.026	-8.644	CIBIO-InBio
DB19704	1 (CZ)	*P. carbonelli*	41.026	-8.645	CIBIO-InBio
DB19703	1 (CZ)	*P. carbonelli*	41.026	-8.644	CIBIO-InBio
DB19610	1 (CZ)	*P. sp*	41.025	-8.645	CIBIO-InBio
DB19725	1 (CZ)	*P. carbonelli*	41.025	-8.644	CIBIO-InBio
DB19662	1 (CZ)	*P. carbonelli*	41.025	-8.645	CIBIO-InBio
DB19833	2 (CZ)	*P. virescens*	40.921	-8.543	CIBIO-InBio
DB19834	2 (CZ)	*P. virescens*	40.921	-8.543	CIBIO-InBio
DB19835	2 (CZ)	*P. carbonelli*	40.921	-8.543	CIBIO-InBio
DB19836	2 (CZ)	*P. carbonelli*	40.921	-8.543	CIBIO-InBio
DB19837	2 (CZ)	*P. carbonelli*	40.921	-8.543	CIBIO-InBio
DB19838	2 (CZ)	*P. carbonelli*	40.921	-8.543	CIBIO-InBio
DB19839	2 (CZ)	*P. carbonelli*	40.921	-8.543	CIBIO-InBio
DB19840	2 (CZ)	*P. carbonelli*	40.921	-8.543	CIBIO-InBio
DB19841	2 (CZ)	*P. carbonelli*	40.921	-8.543	CIBIO-InBio
DB19842	2 (CZ)	*P. carbonelli*	40.921	-8.543	CIBIO-InBio
DB19843	2 (CZ)	*P. carbonelli*	40.921	-8.543	CIBIO-InBio
DB19844	2 (CZ)	*P. virescens*	40.921	-8.543	CIBIO-InBio
DB19845	2 (CZ)	*P. carbonelli*	40.921	-8.543	CIBIO-InBio
DB19846	2 (CZ)	*P. virescens*	40.921	-8.543	CIBIO-InBio
DB19847	2 (CZ)	*P. carbonelli*	40.921	-8.543	CIBIO-InBio
DB19848	2 (CZ)	*P. virescens*	40.921	-8.543	CIBIO-InBio
DB19849	2 (CZ)	*P. virescens*	40.921	-8.543	CIBIO-InBio
DB19850	2 (CZ)	*P. carbonelli*	40.921	-8.543	CIBIO-InBio
DB19851	2 (CZ)	*P. virescens*	40.921	-8.543	CIBIO-InBio
DB19852	2 (CZ)	*P. carbonelli*	40.921	-8.543	CIBIO-InBio
DB19853	2 (CZ)	*P. virescens*	40.921	-8.543	CIBIO-InBio
DB19854	2 (CZ)	*P. virescens*	40.921	-8.543	CIBIO-InBio
DB19855	2 (CZ)	*P. virescens*	40.921	-8.543	CIBIO-InBio
DB19856	2 (CZ)	*P. virescens*	40.921	-8.543	CIBIO-InBio
DB19858	2 (CZ)	*P. carbonelli*	40.921	-8.543	CIBIO-InBio
DB19859	2 (CZ)	*P. carbonelli*	40.921	-8.543	CIBIO-InBio
DB19860	2 (CZ)	*P. carbonelli*	40.921	-8.543	CIBIO-InBio
DB19863	2 (CZ)	*P. carbonelli*	40.921	-8.543	CIBIO-InBio
DB19864	2 (CZ)	*P. carbonelli*	40.921	-8.543	CIBIO-InBio
DB19865	2 (CZ)	*P. carbonelli*	40.921	-8.543	CIBIO-InBio
DB19866	2 (CZ)	*P. carbonelli*	40.921	-8.543	CIBIO-InBio
DB19867	2 (CZ)	*P. carbonelli*	40.921	-8.543	CIBIO-InBio
DB19869	2 (CZ)	*P. virescens*	40.921	-8.543	CIBIO-InBio

Sample label	Contact zone analyzed[1]	Field ID	Latitude	Longitude	Stored at
DB19871	2 (CZ)	*P. virescens*	40.921	-8.543	CIBIO-InBio
DB19873	2 (CZ)	*P. virescens*	40.921	-8.543	CIBIO-InBio
DB19874	2 (CZ)	*P. virescens*	40.921	-8.543	CIBIO-InBio
DB19875	2 (CZ)	*P. virescens*	40.921	-8.543	CIBIO-InBio
DB19876	2 (CZ)	*P. virescens*	40.921	-8.543	CIBIO-InBio
DB19878	2 (CZ)	*P. virescens*	40.921	-8.543	CIBIO-InBio
DB19879	2 (CZ)	*P. virescens*	40.921	-8.543	CIBIO-InBio
DB19880	2 (CZ)	*P. virescens*	40.921	-8.543	CIBIO-InBio
DB19881	2 (CZ)	*P. virescens*	40.921	-8.543	CIBIO-InBio
DB19882	2 (CZ)	*P. virescens*	40.921	-8.543	CIBIO-InBio
DB19883	2 (CZ)	*P. virescens*	40.921	-8.543	CIBIO-InBio
DB19884	2 (CZ)	*P. carbonelli*	40.921	-8.543	CIBIO-InBio
DB19885	2 (CZ)	*P. virescens*	40.921	-8.543	CIBIO-InBio
DB19886	2 (CZ)	*P. virescens*	40.921	-8.543	CIBIO-InBio
DB19887	2 (CZ)	*P. virescens*	40.921	-8.543	CIBIO-InBio
DB19888	2 (CZ)	*P. carbonelli*	40.921	-8.543	CIBIO-InBio
DB19891	3 (CZ)	*P. carbonelli*	40.403	-7.587	CIBIO-InBio
DB19892	3 (CZ)	*P. carbonelli*	40.403	-7.587	CIBIO-InBio
DB19893	3 (CZ)	*P. carbonelli*	40.403	-7.587	CIBIO-InBio
DB19894	3 (CZ)	*P. carbonelli*	40.403	-7.587	CIBIO-InBio
DB19895	3 (CZ)	*P. carbonelli*	40.403	-7.587	CIBIO-InBio
DB19896	3 (CZ)	*P. carbonelli*	40.403	-7.587	CIBIO-InBio
DB19897	3 (CZ)	*P. carbonelli*	40.403	-7.587	CIBIO-InBio
DB19898	3 (CZ)	*P. carbonelli*	40.403	-7.587	CIBIO-InBio
DB19899	3 (CZ)	*P. carbonelli*	40.403	-7.587	CIBIO-InBio
DB19900	3 (CZ)	*P. carbonelli*	40.403	-7.587	CIBIO-InBio
DB19901	3 (CZ)	*P. carbonelli*	40.403	-7.587	CIBIO-InBio
DB21404	3 (CZ)	*P. g. lusitanicus*	40.403	-7.587	CIBIO-InBio
DB21405	3 (CZ)	*P. g. lusitanicus*	40.403	-7.587	CIBIO-InBio
DB21406	3 (CZ)	*P. g. lusitanicus*	40.403	-7.587	CIBIO-InBio
DB21408	3 (CZ)	*P. g. lusitanicus*	40.403	-7.587	CIBIO-InBio
DB21411	3 (CZ)	*P. g. lusitanicus*	40.403	-7.587	CIBIO-InBio
DB21413	3 (CZ)	*P. g. lusitanicus*	40.403	-7.587	CIBIO-InBio
DB21414	3 (CZ)	*P. g. lusitanicus*	40.403	-7.587	CIBIO-InBio
DB21415	3 (CZ)	*P. g. lusitanicus*	40.403	-7.587	CIBIO-InBio
DB21416	3 (CZ)	*P. g. lusitanicus*	40.403	-7.587	CIBIO-InBio
DB21417	3 (CZ)	*P. g. lusitanicus*	40.403	-7.587	CIBIO-InBio
DB21418	3 (CZ)	*P. g. lusitanicus*	40.403	-7.587	CIBIO-InBio
DB21420	3 (CZ)	*P. g. lusitanicus*	40.403	-7.587	CIBIO-InBio

Sample label	Contact zone analyzed[1]	Field ID	Latitude	Longitude	Stored at
DB21421	3 (CZ)	*P. g. lusitanicus*	40.403	-7.587	CIBIO-InBio
DB21423	3 (CZ)	*P. g. lusitanicus*	40.403	-7.587	CIBIO-InBio
DB21424	3 (CZ)	*P. g. lusitanicus*	40.403	-7.587	CIBIO-InBio
DB21432	3 (CZ)	*P. g. lusitanicus*	40.403	-7.587	CIBIO-InBio
DB21434	3 (CZ)	*P. g. lusitanicus*	40.403	-7.587	CIBIO-InBio
DB21435	3 (CZ)	*P. g. lusitanicus*	40.403	-7.587	CIBIO-InBio
DB21436	3 (CZ)	*P. g. lusitanicus*	40.403	-7.587	CIBIO-InBio
DB21471	3 (CZ)	*P. g. lusitanicus*	40.403	-7.587	CIBIO-InBio
DB21472	3 (CZ)	*P. g. lusitanicus*	40.403	-7.587	CIBIO-InBio
DB21474	3 (CZ)	*P. g. lusitanicus*	40.403	-7.587	CIBIO-InBio
DB21475	3 (CZ)	*P. g. lusitanicus*	40.403	-7.587	CIBIO-InBio
DB21476	3 (CZ)	*P. g. lusitanicus*	40.403	-7.587	CIBIO-InBio
DB21477	3 (CZ)	*P. g. lusitanicus*	40.403	-7.587	CIBIO-InBio
DB21478	3 (CZ)	*P. g. lusitanicus*	40.403	-7.587	CIBIO-InBio
DB21479	3 (CZ)	*P. g. lusitanicus*	40.403	-7.587	CIBIO-InBio
DB21480	3 (CZ)	*P. g. lusitanicus*	40.403	-7.587	CIBIO-InBio
DB21482	3 (CZ)	*P. g. lusitanicus*	40.403	-7.587	CIBIO-InBio
DB21483	3 (CZ)	*P. g. lusitanicus*	40.403	-7.587	CIBIO-InBio
DB21484	3 (CZ)	*P. g. lusitanicus*	40.403	-7.587	CIBIO-InBio
DB21485	3 (CZ)	*P. g. lusitanicus*	40.403	-7.587	CIBIO-InBio
DB21486	3 (CZ)	*P. g. lusitanicus*	40.403	-7.587	CIBIO-InBio
DB21487	3 (CZ)	*P. g. lusitanicus*	40.403	-7.587	CIBIO-InBio
DB21488	3 (CZ)	*P. carbonelli*	40.403	-7.587	CIBIO-InBio
DB21491	3 (CZ)	*P. carbonelli*	40.403	-7.587	CIBIO-InBio
DB21492	3 (CZ)	*P. carbonelli*	40.403	-7.587	CIBIO-InBio
DB21493	3 (CZ)	*P. g. lusitanicus*	40.403	-7.587	CIBIO-InBio
DB21494	3 (CZ)	*P. carbonelli*	40.403	-7.587	CIBIO-InBio
DB21495	3 (CZ)	*P. g. Lusitanicus*	40.403	-7.587	CIBIO-InBio
DB21496	3 (CZ)	*P. g. Lusitanicus*	40.403	-7.587	CIBIO-InBio
DB21497	3 (CZ)	*P. carbonelli*	40.403	-7.587	CIBIO-InBio
DB21498	3 (CZ)	*P. carbonelli*	40.403	-7.587	CIBIO-InBio
DB21499	3 (CZ)	*P. g. lusitanicus*	40.403	-7.587	CIBIO-InBio
DB21549	4 (CZ)	*P. vaucheri*	37.003	-6.563	CIBIO-InBio
DB21552	4 (CZ)	*P. vaucheri*	37.003	-6.563	CIBIO-InBio
DB21553	4 (CZ)	*P. vaucheri*	37.003	-6.563	CIBIO-InBio
DB21554	4 (CZ)	*P. vaucheri*	37.003	-6.563	CIBIO-InBio
DB21555	4 (CZ)	*P. vaucheri*	37.003	-6.563	CIBIO-InBio
DB21561	4 (CZ)	*P. carbonelli*	37.003	-6.563	CIBIO-InBio
DB21562	4 (CZ)	*P. carbonelli*	37.003	-6.563	CIBIO-InBio

Sample label	Contact zone analyzed[1]	Field ID	Latitude	Longitude	Stored at
DB21563	4 (CZ)	*P. carbonelli*	37.003	-6.563	CIBIO-InBio
DB21564	4 (CZ)	*P. carbonelli*	37.003	-6.563	CIBIO-InBio
DB21573	4 (CZ)	*P. vaucheri*	37.003	-6.563	CIBIO-InBio
DB21574	4 (CZ)	*P. vaucheri*	37.003	-6.563	CIBIO-InBio
DB21576	4 (CZ)	*P. vaucheri*	37.003	-6.563	CIBIO-InBio
DB21577	4 (CZ)	*P. vaucheri*	37.003	-6.563	CIBIO-InBio
DB21578	4 (CZ)	*P. vaucheri*	37.003	-6.563	CIBIO-InBio
DB21579	4 (CZ)	*P. vaucheri*	37.003	-6.563	CIBIO-InBio
DB21580	4 (CZ)	*P. vaucheri*	37.003	-6.563	CIBIO-InBio
DB21581	4 (CZ)	*P. vaucheri*	37.003	-6.563	CIBIO-InBio
DB21582	4 (CZ)	*P. vaucheri*	37.003	-6.563	CIBIO-InBio
DB21583	4 (CZ)	*P. vaucheri*	37.003	-6.563	CIBIO-InBio
DB21584	4 (CZ)	*P. vaucheri*	37.003	-6.563	CIBIO-InBio
DB21585	4 (CZ)	*P. vaucheri*	37.003	-6.563	CIBIO-InBio
DB21586	4 (CZ)	*P. vaucheri*	37.003	-6.563	CIBIO-InBio
DB21587	4 (CZ)	*P. vaucheri*	37.003	-6.563	CIBIO-InBio
DB21588	4 (CZ)	*P. vaucheri*	37.003	-6.563	CIBIO-InBio
DB21589	4 (CZ)	*P. vaucheri*	37.003	-6.563	CIBIO-InBio
DB21965	4 (CZ)	*P. vaucheri*	37.003	-6.563	CIBIO-InBio
DB21966	4 (CZ)	*P. vaucheri*	37.003	-6.563	CIBIO-InBio
DB21967	4 (CZ)	*P. vaucheri*	37.003	-6.563	CIBIO-InBio
DB21968	4 (CZ)	*P. vaucheri*	37.003	-6.563	CIBIO-InBio
DB21969	4 (CZ)	*P. vaucheri*	37.003	-6.563	CIBIO-InBio
DB21971	4 (CZ)	*P. vaucheri*	37.003	-6.563	CIBIO-InBio
DB21972	4 (CZ)	*P. vaucheri*	37.003	-6.563	CIBIO-InBio
DB21973	4 (CZ)	*P. vaucheri*	37.003	-6.563	CIBIO-InBio
DB21974	4 (CZ)	*P. vaucheri*	37.003	-6.563	CIBIO-InBio
DB21975	4 (CZ)	*P. vaucheri*	37.003	-6.563	CIBIO-InBio
DB21976	4 (CZ)	*P. vaucheri*	37.003	-6.563	CIBIO-InBio
DB21977	4 (CZ)	*P. vaucheri*	37.003	-6.563	CIBIO-InBio
DB21978	4 (CZ)	*P. vaucheri*	37.003	-6.563	CIBIO-InBio
DB21979	4 (CZ)	*P. vaucheri*	37.003	-6.563	CIBIO-InBio
DB21980	4 (CZ)	*P. vaucheri*	37.003	-6.563	CIBIO-InBio
DB21981	4 (CZ)	*P. vaucheri*	37.003	-6.563	CIBIO-InBio
DB21982	4 (CZ)	*P. vaucheri*	37.003	-6.563	CIBIO-InBio
DB21983	4 (CZ)	*P. vaucheri*	37.003	-6.563	CIBIO-InBio
DB21984	4 (CZ)	*P. vaucheri*	37.003	-6.563	CIBIO-InBio
DB21985	4 (CZ)	*P. vaucheri*	37.003	-6.563	CIBIO-InBio
DB21986	4 (CZ)	*P. vaucheri*	37.003	-6.563	CIBIO-InBio

Sample label	Contact zone analyzed[1]	Field ID	Latitude	Longitude	Stored at
DB21987	4 (CZ)	*P. vaucheri*	37.003	-6.563	CIBIO-InBio
DB21988	4 (CZ)	*P. vaucheri*	37.003	-6.563	CIBIO-InBio
DB21989	4 (CZ)	*P. vaucheri*	37.003	-6.563	CIBIO-InBio
DB21990	4 (CZ)	*P. vaucheri*	37.003	-6.563	CIBIO-InBio
DB21991	4 (CZ)	*P. vaucheri*	37.003	-6.563	CIBIO-InBio
DB21992	4 (CZ)	*P. vaucheri*	37.003	-6.563	CIBIO-InBio
DB21993	4 (CZ)	*P. vaucheri*	37.003	-6.563	CIBIO-InBio
DB22074	4 (CZ)	*P. vaucheri*	37.003	-6.563	CIBIO-InBio
DB22076	4 (CZ)	*P. vaucheri*	37.003	-6.563	CIBIO-InBio
DB22077	4 (CZ)	*P. vaucheri*	37.003	-6.563	CIBIO-InBio
DB22081	4 (CZ)	*P. carbonelli*	37.003	-6.563	CIBIO-InBio
DB22082	4 (CZ)	*P. carbonelli*	37.003	-6.563	CIBIO-InBio
DB22083	4 (CZ)	*P. carbonelli*	37.003	-6.563	CIBIO-InBio
DB22085	4 (CZ)	*P. carbonelli*	37.003	-6.563	CIBIO-InBio
DB19902	1 (REF)	*P. bocagei*	41.153	-8.643	CIBIO-InBio
DB19903	1 (REF)	*P. bocagei*	41.153	-8.643	CIBIO-InBio
DB19904	1 (REF)	*P. bocagei*	41.153	-8.643	CIBIO-InBio
DB19905	1 (REF)	*P. bocagei*	41.153	-8.643	CIBIO-InBio
DB19906	1 (REF)	*P. bocagei*	41.153	-8.643	CIBIO-InBio
DB19907	1 (REF)	*P. bocagei*	41.153	-8.643	CIBIO-InBio
DB19908	1 (REF)	*P. bocagei*	41.153	-8.643	CIBIO-InBio
DB19909	1 (REF)	*P. bocagei*	41.153	-8.643	CIBIO-InBio
DB19910	1 (REF)	*P. bocagei*	41.153	-8.643	CIBIO-InBio
DB19911	1 (REF)	*P. bocagei*	41.153	-8.643	CIBIO-InBio
DB19912	1 (REF)	*P. bocagei*	41.153	-8.643	CIBIO-InBio
DB19913	1 (REF)	*P. bocagei*	41.153	-8.643	CIBIO-InBio
DB19914	1 (REF)	*P. bocagei*	41.153	-8.643	CIBIO-InBio
DB19915	1 (REF)	*P. bocagei*	41.153	-8.643	CIBIO-InBio
DB19916	1 (REF)	*P. bocagei*	41.153	-8.643	CIBIO-InBio
DB19917	1 (REF)	*P. bocagei*	41.153	-8.643	CIBIO-InBio
DB19918	1 (REF)	*P. bocagei*	41.153	-8.643	CIBIO-InBio
DB19919	1 (REF)	*P. bocagei*	41.153	-8.643	CIBIO-InBio
DB19920	1 (REF)	*P. bocagei*	41.153	-8.643	CIBIO-InBio
DB19923	1 (REF)	*P. bocagei*	41.153	-8.643	CIBIO-InBio
DB19618	1 (REF)	*P. carbonelli*	40.766	-8.710	CIBIO-InBio
DB19619	1 (REF)	*P. carbonelli*	40.766	-8.710	CIBIO-InBio
DB19620	1 (REF)	*P. carbonelli*	40.766	-8.710	CIBIO-InBio
DB19621	1 (REF)	*P. carbonelli*	40.766	-8.710	CIBIO-InBio
DB19622	1 (REF)	*P. carbonelli*	40.766	-8.710	CIBIO-InBio

Sample label	Contact zone analyzed[1]	Field ID	Latitude	Longitude	Stored at
DB19624	1 (REF)	*P. carbonelli*	40.766	-8.710	CIBIO-InBio
DB19625	1 (REF)	*P. carbonelli*	40.766	-8.710	CIBIO-InBio
DB19626	1 (REF)	*P. carbonelli*	40.766	-8.710	CIBIO-InBio
DB19627	1 (REF)	*P. carbonelli*	40.766	-8.710	CIBIO-InBio
DB19628	1 (REF)	*P. carbonelli*	40.766	-8.710	CIBIO-InBio
DB19629	1 (REF)	*P. carbonelli*	40.766	-8.710	CIBIO-InBio
DB19630	1 (REF)	*P. carbonelli*	40.766	-8.710	CIBIO-InBio
DB19631	1 (REF)	*P. carbonelli*	40.766	-8.710	CIBIO-InBio
DB19632	1 (REF)	*P. carbonelli*	40.766	-8.710	CIBIO-InBio
DB19633	1 (REF)	*P. carbonelli*	40.766	-8.710	CIBIO-InBio
DB19639	1 (REF)	*P. carbonelli*	40.766	-8.710	CIBIO-InBio
DB19640	1 (REF)	*P. carbonelli*	40.766	-8.710	CIBIO-InBio
DB19641	1 (REF)	*P. carbonelli*	40.766	-8.710	CIBIO-InBio
DB19642	1 (REF)	*P. carbonelli*	40.766	-8.710	CIBIO-InBio
DB19634	1, 4 (REF)	*P. carbonelli*	40.766	-8.710	CIBIO-InBio
DB19635	1, 4 (REF)	*P. carbonelli*	40.766	-8.710	CIBIO-InBio
DB19636	1, 4 (REF)	*P. carbonelli*	40.766	-8.710	CIBIO-InBio
DB19638	1, 4 (REF)	*P. carbonelli*	40.766	-8.710	CIBIO-InBio
DB9750	2 (REF)	*P. virescens*	39.831	-8.878	CIBIO-InBio
DB9763	2 (REF)	*P. virescens*	39.831	-8.878	CIBIO-InBio
AK6.154	2 (REF)	*P. virescens*	39.006	-4.006	CIBIO-InBio
AK6.156	2 (REF)	*P. virescens*	39.006	-4.006	CIBIO-InBio
AK6.157	2 (REF)	*P. virescens*	39.006	-4.006	CIBIO-InBio
AK6.62	2 (REF)	*P. virescens*	38.018	-7.864	CIBIO-InBio
AK6.63	2 (REF)	*P. virescens*	38.018	-7.864	CIBIO-InBio
AK6.120	2 (REF)	*P. virescens*	38.004	-6.008	CIBIO-InBio
AK6.121	2 (REF)	*P. virescens*	38.004	-6.008	CIBIO-InBio
AK6.123	2 (REF)	*P. virescens*	38.004	-6.008	CIBIO-InBio
AK6.124	2 (REF)	*P. virescens*	38.004	-6.008	CIBIO-InBio
AK6.134	2 (REF)	*P. virescens*	37.874	-3.214	CIBIO-InBio
DB22134	2, 3 (REF)	*P. carbonelli*	37.048	-6.567	CIBIO-InBio
DB22135	2, 3 (REF)	*P. carbonelli*	37.048	-6.567	CIBIO-InBio
DB8245	2, 3, 4 (REF)	*P. carbonelli*	Unk. S. Pedro do Sul, Portugal		CIBIO-InBio
DB8284	2, 3, 4 (REF)	*P. carbonelli*	Unk. Meco, Portugal		CIBIO-InBio
AK4.42	2, 3, 4 (REF)	*P. carbonelli*	Unk. Berlengas, Portugal		CIBIO-InBio
DB9638	2, 3, 4 (REF)	*P. carbonelli*	39.755	-9.032	CIBIO-InBio
DB9639	2, 3, 4 (REF)	*P. carbonelli*	39.755	-9.032	CIBIO-InBio
AK4.142	2, 3, 4 (REF)	*P. carbonelli*	38.709	-9.485	CIBIO-InBio
AK4.143	2, 3, 4 (REF)	*P. carbonelli*	38.709	-9.485	CIBIO-InBio

Sample label	Contact zone analyzed[1]	Field ID	Latitude	Longitude	Stored at
AK4.178	2, 3, 4 (REF)	*P. carbonelli*	37.340	-8.854	CIBIO-InBio
AK4.181	2, 3, 4 (REF)	*P. carbonelli*	37.340	-8.854	CIBIO-InBio
DB21569	2, 3, 4 (REF)	*P. carbonelli*	37.048	-6.567	CIBIO-InBio
DB22136	3 (REF)	*P. carbonelli*	37.048	-6.567	CIBIO-InBio
AK5.146	3 (REF)	*P. g. lusitanicus*	Unk. Isla de Rua, Galicia, Spain		CIBIO-InBio
AK5.147	3 (REF)	*P. g. lusitanicus*	Unk. Isla de Rua, Galicia, Spain		CIBIO-InBio
AK5.148	3 (REF)	*P. g. lusitanicus*	Unk. Isla de Rua, Galicia, Spain		CIBIO-InBio
AK5.149	3 (REF)	*P. g. lusitanicus*	Unk. Isla de Rua, Galicia, Spain		CIBIO-InBio
AK5.150	3 (REF)	*P. g. lusitanicus*	Unk. Isla de Rua, Galicia, Spain		CIBIO-InBio
DB22618	3 (REF)	*P. g. lusitanicus*	50.667	-8.327	CIBIO-InBio
DB22619	3 (REF)	*P. g. lusitanicus*	50.667	-8.327	CIBIO-InBio
DB22620	3 (REF)	*P. g. lusitanicus*	50.667	-8.327	CIBIO-InBio
BEV.2052	3 (REF)	*P. g. lusitanicus*	42.706	-6.630	EPHE-CEFE
BEV.8346	3 (REF)	*P. g. lusitanicus*	42.609	-6.259	EPHE-CEFE
BEV.8333	3 (REF)	*P. g. lusitanicus*	42.038	-6.620	EPHE-CEFE
BEV.8334	3 (REF)	*P. g. lusitanicus*	42.038	-6.620	EPHE-CEFE
AK5.262	3 (REF)	*P. g. lusitanicus*	42.032	-6.268	CIBIO-InBio
AK5.263	3 (REF)	*P. g. lusitanicus*	42.032	-6.268	CIBIO-InBio
AK5.264	3 (REF)	*P. g. lusitanicus*	42.032	-6.268	CIBIO-InBio
DB22612	3 (REF)	*P. g. lusitanicus*	40.702	-8.359	CIBIO-InBio
DB22613	3 (REF)	*P. g. lusitanicus*	40.702	-8.359	CIBIO-InBio
DB22616	3 (REF)	*P. g. lusitanicus*	40.702	-8.359	CIBIO-InBio
DB22617	3 (REF)	*P. g. lusitanicus*	40.702	-8.359	CIBIO-InBio
DB22654	3 (REF)	*P. g. lusitanicus*	40.351	-7.095	CIBIO-InBio
DB22655	3 (REF)	*P. g. lusitanicus*	40.351	-7.095	CIBIO-InBio
DB22656	3 (REF)	*P. g. lusitanicus*	40.351	-7.095	CIBIO-InBio
DB22657	3 (REF)	*P. g. lusitanicus*	40.351	-7.095	CIBIO-InBio
DB22658	3 (REF)	*P. g. lusitanicus*	40.351	-7.095	CIBIO-InBio
DB22659	3 (REF)	*P. g. lusitanicus*	40.351	-7.095	CIBIO-InBio
DB22660	3 (REF)	*P. g. lusitanicus*	40.351	-7.095	CIBIO-InBio
DB22662	3 (REF)	*P. g. lusitanicus*	40.351	-7.095	CIBIO-InBio
DB22663	3 (REF)	*P. g. lusitanicus*	40.351	-7.095	CIBIO-InBio
DB24411	3 (REF)	*P. g. lusitanicus*	40.351	-7.095	CIBIO-InBio
DB24412	3 (REF)	*P. g. lusitanicus*	40.351	-7.095	CIBIO-InBio
DB24413	3 (REF)	*P. g. lusitanicus*	40.351	-7.095	CIBIO-InBio
DB24414	3 (REF)	*P. g. lusitanicus*	40.351	-7.095	CIBIO-InBio
DB24415	3 (REF)	*P. g. lusitanicus*	40.351	-7.095	CIBIO-InBio
DB24424	3 (REF)	*P. g. lusitanicus*	40.351	-7.095	CIBIO-InBio
DB24425	3 (REF)	*P. g. lusitanicus*	40.351	-7.095	CIBIO-InBio

Sample label	Contact zone analyzed[1]	Field ID	Latitude	Longitude	Stored at
DB24426	3 (REF)	*P. g. lusitanicus*	40.351	-7.095	CIBIO-InBio
DB24428	3 (REF)	*P. g. lusitanicus*	40.351	-7.095	CIBIO-InBio
DB8222	3, 4 (REF)	*P. carbonelli*	Unk. Playa de Rompeculos, Andalucia, Spain		CIBIO-InBio
DB8402	3, 4 (REF)	*P. carbonelli*	Unk. São Pedro do Sul, Portugal		CIBIO-InBio
DB22128	3, 4 (REF)	*P. carbonelli*	37.048	-6.567	CIBIO-InBio
DB19085	3,4 (REF)	*P. carbonelli*	Unk. Rompeculos, Andalucia, Spain		CIBIO-InBio
DB9637	4 (REF)	*P. carbonelli*	39.755	-9.032	CIBIO-InBio
AK8.55	4 (REF)	*P. vaucheri SSp*	Unk. Torcal de Antequera, Andalucia, Spain		CIBIO-InBio
AK8.56	4 (REF)	*P. vaucheri SSp*	Unk. Torcal de Antequera, Andalucia, Spain		CIBIO-InBio
AK8.60	4 (REF)	*P. vaucheri SSp*	Unk. Playa de la Vibora, Andalucia, Spain		CIBIO-InBio
AK8.61	4 (REF)	*P. vaucheri SSp*	Unk. Playa de la Vibora, Andalucia, Spain		CIBIO-InBio
AK8.62	4 (REF)	*P. vaucheri SSp*	Unk. Playa de la Vibora, Andalucia, Spain		CIBIO-InBio
AK8.63	4 (REF)	*P. vaucheri SSp*	Unk. Playa de la Vibora, Andalucia, Spain		CIBIO-InBio
AK8.88	4 (REF)	*P. vaucheri SSp*	Unk. Alcala la Real, Andalucia, Spain		CIBIO-InBio
AK8.89	4 (REF)	*P. vaucheri SSp*	Unk. Alcala la Real, Andalucia, Spain		CIBIO-InBio
BEV.12849	4 (REF)	*P. vaucheri SSp*	37.932	-4.663	EPHE-CEFE
AK8.112	4 (REF)	*P. vaucheri SSp*	37.786	-2.225	CIBIO-InBio
AK8.113	4 (REF)	*P. vaucheri SSp*	37.786	-2.225	CIBIO-InBio
DB11186	4 (REF)	*P. vaucheri SSp*	36.373	-6.187	CIBIO-InBio
DB11187	4 (REF)	*P. vaucheri SSp*	36.373	-6.187	CIBIO-InBio
DB11198	4 (REF)	*P. vaucheri SSp*	36.373	-6.187	CIBIO-InBio
DB11255	4 (REF)	*P. vaucheri SSp*	36.373	-6.187	CIBIO-InBio

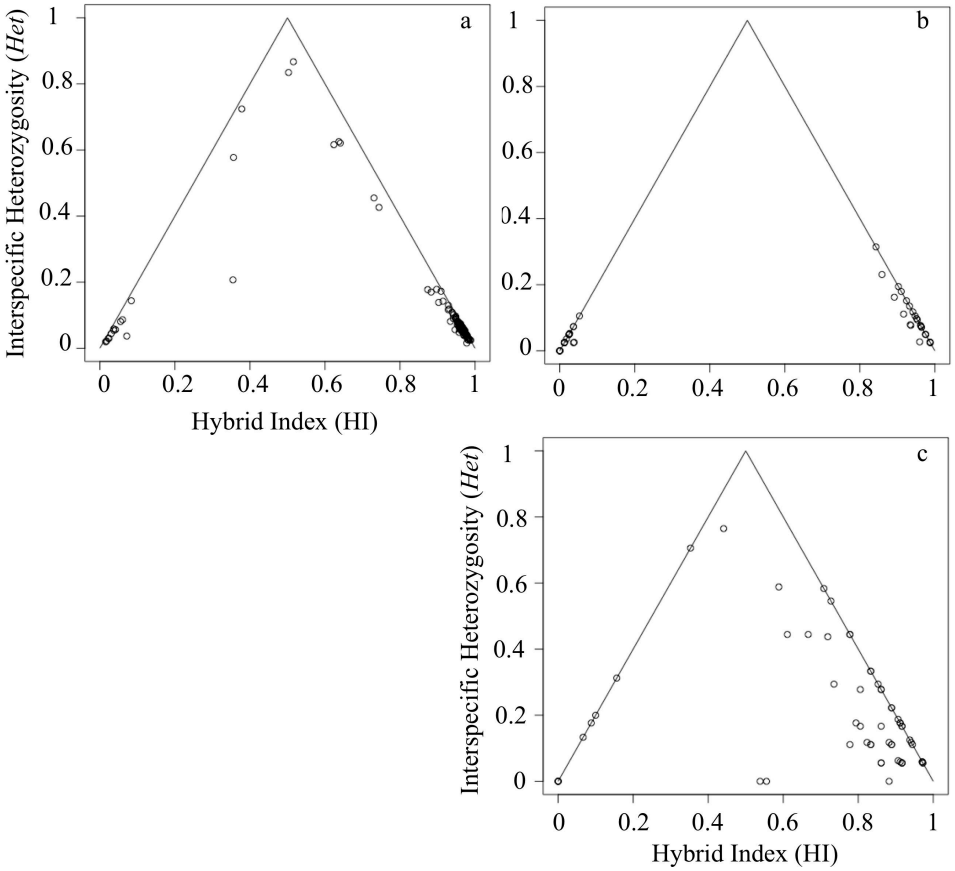

Figure S3.1. Distributions of individual hybrid index and interspecific heterozygosity in the contact zones between **a.** *P. bocagei* x *P. carbonelli*, **b.** *P. virescens* x *P. carbonelli*, **c.** *P. vaucheri* SSp x *P. carbonelli* based on diagnostic loci. Results for *P. guadarramae lusitanicus* x *P. carbonelli* contact zone are missing because we found only one diagnostic loci among the 1085 discovered for this contact zone. In each contact zone *P. carbonelli* reference individuals were set to have a HI of 0, and the other species were set to an HI of 1.

Appendix IV. A case of *Podarcis carbonelli* intake by *Podarcis virescens*

Guilherme Dias[1,2,3,*], Carla Luis[1], Catarina Pinho[1] and Antigoni Kaliontzopoulou[1,4]

[1]CIBIO-InBio, Centro de Investigação em Biodiversidade e Recursos Genéticos, Portugal. [*]Corresponding author. E-mail: guilhermedias@cibio.up.pt

[2]Departamento de Biologia, Faculdade de Ciências da Universidade do Porto, Portugal [3] CEFE, [3]CEFE, CNRS, University of Montpellier, University Paul Valéry Montpellier 3, EPHE, IRD, Montpellier, France

[4]Department of Ecology, Evolution, and Organismal Biology, Iowa State University, Ames, IA 50011, USA

Podarcis virescens (*P. hispanica* type 2) and *P. carbonelli* are two of the evolutionary lineages that constitute the *Podarcis hispanica* species complex (Kaliontzopoulou et al. 2011) and both are endemic to the Iberian Peninsula. *P. virescens* (*P. hispanica* type 2) is distributed in Central and Southern Iberian Peninsula, except for the most southern, southeastern and eastern extremes of Spain (Sá-Sousa et al. 2002, Kaliontzopoulou et al. 2011, Geniez et al. 2014). *P. carbonelli* has a very fragmented distribution, occupying the Northern-Central region of Portugal between Douro and Mondego rivers, the west Central System from Serra da Estrela to Sierra de Francia, the west Portuguese coast south to Douro river and an isolated population at Doñana National Park at southwestern coast of Spain (Sá-Sousa 2000, 2004, 2008, Kaliontzopoulou et al. 2011, Sillero and Carretero 2013). As most small lacertids (see Arnold 1987, Diaz and Carrascal 1993) these two lizards mostly prey on small arthropods as *Coleoptera*, *Homoptera* and *Aranea* (Pérez-Mellado 1998a, 1998b). Among several biotic and abiotic factors determining feeding behaviour in lacertids (Arnold 1987, Carretero 2004), intraspecific predation, i.e. cannibalism, is frequently attributed to high densities and reduced resource availability characteristic of insular populations (e.g. Pérez-Mellado and Corti 1993). Usually, the cannibal predator is considerably larger than the prey (Polis 1981), which are most frequently juveniles. Intraspecific predation is a phenomenon widespread across several vertebrate groups as fishes (Smith and Reay 1991), amphibians, and reptiles (Polis and Myers 1985). Also predation events between closely related species, including congenerics, have been reported in fishes (e.g. *Onchorhynchus* spp. Taniguchi et al. 2002), amphibians (e.g. *Ambystoma* spp. Stenhouse et al. 1983, Cortwright 1988), and reptiles (e.g. *Anolis* spp. Stamps 1983; Gerber and Echternacht 2000). Furthermore, this type of predation frequently occurs between species with similar ecological requirements exploiting the same resources – a phenomenon termed intraguild predation – where individuals prey over direct competitors (Polis et al. 1989). Several cases of intraspecific predation have been reported for insular populations of some species of the lizard genus *Podarcis*. For instance, adults of *P. atrata* from Columbretes Islands are known to predate on both conspecific eggs and juveniles, where males have been shown to have a higher propensity towards cannibalism than females (Castilla 1995, Castilla and Van Damme 1996). Similarly, some conspecific juveniles were found in the diet of both males and females of *P. filfolensis* from Linosa (Bombi et al. 2005) and Lampione (Carretero et al. 2010) islands in the Pelagian Archipelago, as well as of *P.*

gaigeae from Skyros (Adamopoulou et al. 1999). Although less frequent, cannibalism has also been reported for continental populations of *Podarcis*, e.g. for *P. muralis* from Kabischki in east Bulgaria (Engelmann 1964 *in* Polis and Myers 1985) and from Vransko in central Slovenia, where this species is found in populations with high local densities (Žagar and Carretero 2012). Adult males of *P. sicula* have also been reported to pray over juveniles of their own species, e.g. in central (Rugiero 1994, Grano et al. 2011) and southern Italy (Capula and Aloise 2011) and in an introduced population in New York (Burke and Mercurio 2002). Interestingly, despite the fact that various areas of distribution overlap and even strict syntopy among different *Podarcis* species exist (Gasc et al. 1997, Arnold and Ovenden 2002), only one case of interspecific predation attempt has been reported, to our knowledge. Remarkably, this single observation comes from an introduced population of an insular species, *P. pityusensis*, in Barcelona, where an adult was seen attacking, but not eating, an autochthonous *P. liolepis* male (Carretero and Llorente 2001). While sampling in Santa Maria da Feira (40.92° N, 8.54° E, Datum WGS1984; Figure S4.1), Aveiro, Portugal, we captured a *P. virescens* male that regurgitated a *Podarcis* tail, while being manipulated immediately after the capture. The sampling was carried out on July 9, 2013, in the medieval castle of Santa Maria da Feira where both species occur in strict syntopy. The captured lizard was 58 mm in snout vent length (SVL). The regurgitated tail was small, fresh, non-regenerated and with small segments indicating that it belonged to a newborn lizard. To confirm the morphological identification of the predator, and identify the species of the prey, we used genetic tools because the specific identification is impossible based on tail morphology only. Genomic DNA was extracted from alcohol-preserved tail muscle following the EasySpin® Genomic DNA Tissue Kit manufacturer's instructions. For both individuals we sequenced a 16S rRNA gene fragment (504 bp) with the primers 16sL1 and 16sH1 as described in Hedges and Bezy (1993) and a ND4-tRNALEU gene fragment (809 bp) with the primers ND4 and Leu as described in Arévalo et al. (1994). PCR products were sequenced with the same primers used for amplification of ND4- tRNALEU and with 16sL1 for 16S rRNA following the ABI PRISM BigDye Terminator Cycle Sequencing 3.1 (Applied Biosystems) standard protocol. These sequences were edited by hand with BioEdit version 7.2.5 (Hall 1999) then compared to sequences available in GenBank® using the BLAST® tool. Genetic analysis corroborates the morphological identification of the captured lizard as a *P. virescens* and also validates the regurgitated tail as belonging to a *P. carbonelli*

individual. The 16s gene fragments match 99% with other 16s fragments from both *P. virescens* (accession numbers DQ081081 – DQ081084) and *P. carbonelli* (accession numbers DQ081081 – DQ081084). Similar results were obtained with the ND4 gene fragments that match 99% with other *P. carbonelli* fragments of the same gene (accession numbers EF081135 – EF081156 and DQ081154 – DQ081155) and 99% with *P. virescens* (accession numbers DQ081159, EU269563 and EU269567). Note that *P. virescens* and *P. carbonelli* sequences are different by 4.37% and 8.04% on average for the 16S and ND4 fragments analyzed here, meaning that the genetic identification is straightforward. Sequences are deposited in GenBank® under the accession numbers KP455498 and KP455500 for *P. virescens* and KP455499 and KP455501 for *P. carbonelli* respectively. Because we could only recover the tail of the *P. carbonelli* juvenile, we cannot be sure if the captured lizard consumed the whole animal or just the autotomized tail. In any case, some characteristics of the observed event are worth noting. The case reported here consists of a single observation, and we cannot reach general conclusions on its frequency, but the marked difference in size between the predator *P. virescens* (mean SVL = 53.23; Kaliontzopoulou 2010) and prey *P. carbonelli* (mean SVL = 50.20; Kaliontzopoulou 2010) reported here fits previous observations, as several cases of cannibalism are known in *Podarcis* species, where large adults, mainly males, attack and/or prey over juveniles (Rugiero 1994, Castilla 1995, Castilla and Van Damme 1996, Burke and Mercurio 2002, Capula and Aloise 2011, Grano et al. 2011, Žagar and Carretero 2012). The time of the year at which the sampling was carried out coincides with the birth of new hatchlings, which seems to be the case of the *P. carbonelli*, and the differences in size between large adults and newborns are very marked. The consumed prey would certainly represent an important energy intake for the predator (Polis et al. 1989). In addition, elevated lizard density also seems to be a common factor associated to observations of cannibalism and predatory attacks towards co-specifics (Pérez-Mellado and Corti 1993), a case not restricted to insular populations (e.g. Žagar and Carretero 2012). In the case reported here, both species are strictly syntopic and relatively abundant at a local scale, increasing interspecific encounters and yielding antagonistic interactions more prone to occur. In cases of co-occurrence of closely related species, intraguild predation could act as a community structure regulator, where juveniles are particularly vulnerable (Polis et al. 1989, Polis and Holt 1992). Therefore it is not surprising that a large adult male would predate over a juvenile

from a competitor species. Both necrofagia and the opportunistic consumption of the autotomized tail are events very unlikely to occur. Nevertheless as we cannot assess the attempt of predation, we cannot completely discard both hypotheses. The lack of known interspecific predation events among *Podarcis* species may be associated to a low frequency of occurrence of such events in natural populations. Additionally, their incidence is expected to be higher in some temporal and spatial frames, as during the hatching period and in high density populations. Since this kind of biological interactions represent a high energy intake to the predator and could interfere in the community structure, they represent an intriguing field for future investigation to better understand its ecological importance among lizards.

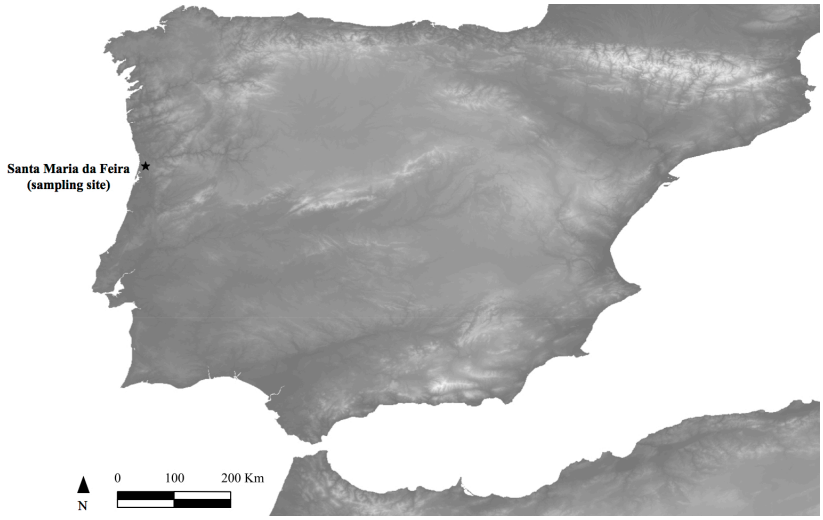

Figure S4.1. Iberian Peninsula map with the geographic location of the syntopyc *Podarcis virescens* and *P. carbonelli* populations from Santa Maria da Feira (Portugal).

Acknowledgments

GD was supported by a PhD grant (SFRH/BD/89750/2012), CL by a research grant (PTDC/ BIA-BEC/102179/2008), and CP and AK by post-doctoral grants (SFRH/BPD//87204/2012 and SFRH/BPD/68493/2010 respectively), all under the Programa Operacional Potencial Humano – Quadro de Referência Estratégico Nacional funds from the European Social Fund and Portuguese Ministério da Educação e Ciência. Support was also provided by FCT project PTDC/BIA-BEC/102179/2008, under FEDER COMPETE funds (FCOMP-01-0124-FEDER-007062) and by the project "Biodiversity, Ecology and Global Change" cofinanced by North Portugal Regional Operational Programme 2007/2013 (ON.2 – O Novo Norte), under the National Strategic Reference Framework (NSRF), through the European Regional Development Fund (ERDF).

References

Adamopoulou, C., Valakos, E.D., Pafilis, P. (1999): Summer diet of *Podarcis milensis*, *P. gaigeae* and *P. erhardii* (Sauria: Lacertidae). Bonn. Zool. Beitr. 48: 275-282.

Arévalo, E., Davis, S.K., Sites, J.W. (1994): Mitochondrial DNA sequence divergence and phylogenetic relationships among eight chromosome races of the *Sceloporus grammicus* complex (Phrynosomatidae) in central Mexico. Syst. Biol. 43: 387–418. Arnold, E.N. (1987): Resource partition among lacertid lizards in southern Europe. J. Zool. 1: 739-782.

Arnold, E.N., Ovenden, D.W. (2002): A field guide to the reptiles and amphibians of Britain and Europe. Harper Collins Publisher, London.

Bombi, P., Vignoli, L., Scalera, R., Bologna, M.A. (2005): Food habits of *Podarcis filfolensis* (Reptilia, Lacertidae) on a small Mediterranean island during the dry season. Amphibia-Reptilia 26: 412-417.

Burke, R.L., Mercurio, R.J. (2002): Food Habits of a New York Population of Italian Wall Lizards, *Podarcis sicula* (Reptilia, Lacertidae). Am. Midl. Nat. 147: 368-375.

Carretero, M.A. (2004): From set menu to *a la carte*. Linking issues in trophic ecology of Mediterranean lacertids. Ital. J. of Zool. 2: 121-133.

Carretero, M.A., Lo Cascio, P., Corti, C., Pasta, S. (2010): Sharing resources in a tiny Mediterranean island? Comparative diets of Chalcides ocellatus and *Podarcis filfolensis* in Lampione. Bonn. Zool. Beitr. 57: 111-118.

Carretero, M.A., Llorente, G.A. (2001): What are they really eating? Stomach *versus* intestine as sources of diet information in lacertids. pp. 105-112, *in*: Vicente L and Crespo EG (eds.) *Mediterranean Basin Lacertid Lizards. A Biological Approach*. ICN, Lisboa, Portugal.

Castilla, A.M. (1995): Conspecific eggs and hatchlings in the diet of the insular lizard *Podarcis hispanica atrata*. Bolletín Soc. Hist. Nat. Balears 38: 121-129.

Castilla, A.M., Van Damme, R. (1996): Cannibalistic Propensities in the Lizard *Podarcis hispanica* atrata. Copeia 4: 991-994.

Capula, M., Aloise, G. (2011): Extreme feeding behaviours in the Italian wall lizard, *Podarcis siculus*. *Acta Herpetologica* 6: 11- 14.

Cortwright, S.A. (1988): Intraguild predation and competition: an analysis of net growth shifts in larval amphibian prey. Can. J. Zool. 66: 1813-1813.

Diaz, J.A., Carrascal, L.M. (1993): Variation in the effect of profitability on prey size selection by the lacertid lizard Psammodromus algirus. Oecologia 94: 23-29.

Gasc, J-P., Cabela, A., Crnobrnja-Isailovic, J., Dolmen, D., Grossenbacher, K., Haffner, P., Lescure, J., Martens, H., Martínez Rica, J.P., Maurin, H., Oliveira, M.E., Sofianidou T.S., Veith, M., Zuiderwijk, A. (1997) Atlas of amphibians and reptiles in Europe. Societas Europaea Herpetologica and Muséum National d'Histoire Naturelle, Paris.

Geniez, P., Sá-Sousa, P., Guillaume, C.P., Cluchier A., Crochet, P-A. (2014) Systematics of the *Podarcis hispanicus* complex (Sauria, Lacertidae) III: valid nomina of the western and central Iberian forms. *Zootaxa* 3794: 001-051.

Gerber, G.P., Echternacht, A.C. (2000): Evidence for asymmetrical intraguild predation between native and introduced *Anolis* lizards. Oecologica 124: 599-607.

Grano, M., Cattaneo, C., Cattaneo, A. (2011): A case of cannibalism in P*odarcis siculus campestris* De Betta, 1857 (Reptilia, Lacertidae). Biodiversity Journal 2: 151-152.

Hall T.A. (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95／98／NT. Nucleic Acids Symposium Series, 41, 95–98.

Hedges, S.B., Bezy R.L. (1993) Phylogeny of Xantusiid lizards: concern for data and analysis. Mol. Phylogenet. Evol. 2: 76–87.

Kaliontzopoulou, A. (2010): Proximate and evolutionary causes of phenotypic diversification: morphological variation in Iberian and North African *Podarcis* wall lizards. PhD Thesis, University of Barcelona. 227pp.

Kaliontzopoulou, A., Pinho, C., Harris, D.J, Carretero, M.A. (2011): When cryptic diversity blurs the picture: a cautionary tale from Iberian and North African Podarcis wall lizards. Biol. J. Linn. Soc. 103: 779–800.

Pérez-Mellado, V. (1998a): *Podarcis bocagei*, *in*: Salvador A (Coord) Reptiles, Ramos et al. (Eds.) Fauna Ibérica, vol 10, segunda impression, Museo Nacional de Ciencias Naturales, CSIC, Madrid, pp. 243-257.

Pérez-Mellado, V. (1998b): *Podarcis hispanica*, *in*: Salvador A (Coord) Reptiles, Ramos et al. (Eds.) FaunaIbérica,vol10, segunda impression, Museo Nacional de Ciencias Naturales, CSIC, Madrid, pp. 258-272.

Pérez-Mellado, V., Corti, C. (1993): Dietary adaptations and herbivory in lacertid lizards of the genus *Podarcis* from western Mediterranean islands (Reptilia: Sauria). Bonn. Zool. Beitr. 44: 193-220.

Polis, G.A. (1981): The evolution and dynamics of intraspecific predation. Annu. Rev. Ecol. Syst. 12: 225-251.

Polis, G.A., Holt, R.D. (1992): Intraguild predation: the dynamics of complex trophic interactions. Trends Ecol. Evol. 7: 151-154.

Polis, G.A., Myers, C.A. (1985): A survey of intraspecific predation among reptiles and amphibians. J. Herpetol. 19: 99-107.

Polis, G.A., Myers, C.A., Holt, R.D. (1989): The ecology and evolution of intraguild predation: potential competitor that eat each other. Annu. Rev. Ecol. Evol. Syst. 20: 297-330.

Rugiero, L. (1994): Food habits of the Ruin Lizard, *Podarcis sicula* (Rafinesque-Schmaltz, 1810), from a coastal dune in Central Italy (Squamata: Sauria: Lacertidae). Herpetozoa 7: 71-73.

Sá-Sousa, P. (2000): Distribución de la lagartija *Podarcis carbonelli* (Pérez-Mellado, 1981) en Portugal. Bol. Asoc. Herpetol. Esp. 11: 12-15.

Sá-Sousa, P. (2004): *Podarcis carbonelli, in*: Pleguezuelos JM, Márquez R, Lizana M (eds.), *Atlas y Libro Rojo de los Anfibios y Reptiles de España*. Direccíon General de Conservación de la Naturaleza – Asociación Herpetológica Española (3ª impresión), Madrid, pp. 243-244.

Sá-Sousa, P. (2008): *Podarcis carbonelli*, *in*: Loureiro A, Ferrand de Almeida N, Carretero MA, Paulo OS (eds.), *Atlas dos Anfibios e Répteis de Portugal*. Instituto da Conservação da Natureza, Lisboa, pp. 152-153.

Sá-Sousa, P., Vicente, L., Crespo E.G. (2002): Morphological variability of *Podarcis hispanica* (Sauria: Lacertidae) in Portugal. Amphibia-Reptilia 23: 55–69.

Sillero, N., Carretero, M.A. (2013): Modelling the past and future distribution of contracting species. The Iberian lizard *Podarcis carbonelli* (Squamata: Lacertidae) as a case study. Zool. Anz. 252: 289–298.

Smith, C., Reay, P. (1991): Cannibalism in teleost fish. Rev. Fish Biol. Fish. 1: 41-64. Stamps, J. (1983): The relationships between ontogenetic habitat shifts, competition and predator avoidance in a juvenile lizard. (*Anolis aeneus*). Behav. Ecol. Sociobiol. 12: 19-33.

Stenhouse, S.L., Heirston, N.G., Cobey, A.E. (1983): Predation and competition in *Ambystoma* larvae: field and laboratory experiments. J. Herpet. 17: 210-220.

Taniguchi, Y., Fausch, K.D., Nakano, S. (2002): Size-structured Interactions between Native and Introduced Species: Can Intraguild Predation Facilitate Invasion by Stream Salmonids? *Biological Invasions* 4: 223-233.

Žagar, A., Carretero, M.A. (2012): A record of cannibalism in *Podarcis muralis* (Laurenti, 1768) (Reptilia, Lacertidae) from Slovenia. Herpetol. Notes 5: 211-213.